



UNIVERSIDADE DE BRASÍLIA – UnB
FACULDADE DE ECONOMIA, ADMINISTRAÇÃO E CONTABILIDADE – FACE
DEPARTAMENTO DE CIÊNCIAS CONTÁBEIS E ATUARIAIS - CCA
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIAS CONTÁBEIS - PPGCONT

DANIEL DA SILVA MELLO

EFICIÊNCIA DAS INSTITUIÇÕES DE ENSINO SUPERIOR NO BRASIL: Uma
Análise DEA sobre Alocação de Recursos

Brasília - DF

2026

DANIEL DA SILVA MELLO

EFICIÊNCIA DAS INSTITUIÇÕES DE ENSINO SUPERIOR NO BRASIL: Uma
Análise DEA sobre Alocação de Recursos

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Ciências Contábeis (PPGCont) da Faculdade de Economia, Administração e Contabilidade da Universidade de Brasília (UnB) como requisito para a obtenção de título de Mestre em Ciências Contábeis.

Orientador: Prof. Dr. Carlos Rosano-Peña

Brasília - DF

2026

FICHA CATALOGRÁFICA

MELLO, Daniel da Silva

EFICIÊNCIA DAS INSTITUIÇÕES DE ENSINO SUPERIOR NO BRASIL: Uma Análise DEA sobre Alocação de Recursos / Daniel da Silva Mello; orientador Professor Doutor Carlos Rosano-Peña. - Brasília, 2026.

205 p.

Dissertação (Mestrado em Ciências Contábeis) - Universidade de Brasília, 2026.

1. ensino superior. 2. eficiência técnica. 3. DEA em rede. 4. Fies. 5. políticas públicas. I. Peña, Professor Doutor Carlos Rosano-Pena, orient. II. Título.

Professora Doutora Rozana Reigota Naves
Reitora

Professor Doutor Márcio Muniz de Farias
Vice-Reitor

Professor Doutor Roberto Goulart Menezes
Decano de Pós-graduação

Professor Doutor Roberto de Góes Ellery Júnior
**Diretor da Faculdade de Economia, Administração, Contabilidade e Gestão de
Políticas Públicas**

Professor Doutor Wagner Rodrigues dos Santos
Chefe do Departamento de Ciências Contábeis e Atuariais

Professor Doutor César Augusto Tibúrcio Silva
Coordenador do Programa de Pós-graduação em Ciências Contábeis da UnB

DANIEL DA SILVA MELLO

EFICIÊNCIA DAS INSTITUIÇÕES DE ENSINO SUPERIOR NO BRASIL: Uma
Análise DEA sobre Alocação de Recursos

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Ciências Contábeis (PPGCont) da Faculdade de Economia, Administração e Contabilidade da Universidade de Brasília (UnB) como requisito para a obtenção de título de Mestre em Ciências Contábeis.

BANCA EXAMINADORA:

Prof. Dr. Carlos Rosano-Peña

Presidente – Universidade de Brasília (PPGCont/UnB)

Prof. Dr. Abimael de Jesus Barros Costa

Examinador Interno – Universidade de Brasília (PPGCont/UnB)

Profa. Dra. Carla Velloso dos Santos

Examinadora Externa – Universidade Federal do Rio de Janeiro (PPGCC/UFRJ)

Prof. Dr. Marcelo Driemeyer Wilbert

Suplente Interno – Universidade de Brasília (PPGCont/UnB)

Brasília, 27 de fevereiro de 2026.

AGRADECIMENTOS

A Deus, por ter me concedido saúde, força e proteção em todos os momentos desta caminhada, especialmente nos períodos de maior desafio e dedicação necessários à conclusão deste mestrado.

À minha esposa, Monique Bastos Breder, pela compreensão, incentivo e apoio incondicional. Sua companhia tornou mais leves as angústias e o cansaço gerados pelo acúmulo de responsabilidades profissionais, pessoais e acadêmicas. Obrigado por acreditar neste projeto desde o início e por estar sempre ao meu lado.

Aos meus pais, Dorival de Mello Filho e Mariza da Silva Mello, por me ensinarem desde cedo o valor transformador da educação e por nunca medirem esforços em me incentivar a seguir em frente, mesmo diante das dificuldades. Vocês me mostraram que a aprendizagem é capaz de mudar destinos e abrir horizontes.

Aos meus filhos, Beatriz de Mello Breder e Bruno de Mello Breder, razão da minha maior inspiração e força. Que este trabalho seja um exemplo de que com dedicação, disciplina e perseverança é possível superar desafios e conquistar novos objetivos. Que vocês sigam sempre acreditando no poder transformador do conhecimento e da educação.

Ao meu orientador, Prof. Dr. Carlos Rosano-Peña, pela orientação cuidadosa, paciência e constante dedicação. Sua confiança e generosidade foram fundamentais para o desenvolvimento desta pesquisa e para o meu amadurecimento acadêmico.

À Profa. Dra. Carla Velloso dos Santos, da Universidade Federal do Rio de Janeiro (PPGCC/UFRJ), e ao Prof. Dr. Marcelo Driemeyer Wilbert, da Universidade de Brasília (PPGCont/UnB), pelas valiosas contribuições acadêmicas na etapa de qualificação e defesa, que se refletiram diretamente na lapidação e no aprimoramento deste trabalho. Suas observações críticas, sugestões e questionamentos foram fundamentais para elevar a qualidade desta dissertação.

Aos professores que marcaram minha trajetória acadêmica no mestrado, transmitindo ensinamentos que moldaram minha formação e ampliaram minha visão crítica. Em especial, ao Prof. Dr. Abimael de Jesus Barros Costa, por acreditar no meu potencial, incentivar a continuidade da carreira acadêmica e, sobretudo, por se constituir em um grande exemplo de educador e pesquisador a ser seguido.

Por fim, agradeço a todos que, de forma direta ou indireta, contribuíram para a realização deste trabalho e para o êxito desta jornada.

RESUMO

MELLO, Daniel da Silva. *EFICIÊNCIA DAS INSTITUIÇÕES DE ENSINO SUPERIOR NO BRASIL: Uma Análise DEA sobre Alocação de Recursos*. 2026. 205f. Dissertação (Mestrado em Ciências Contábeis) – Universidade de Brasília, Brasília, 2026.

O presente estudo teve como objetivo avaliar a eficiência técnica aplicada ao processo produtivo das Instituições de Ensino Superior (IES) brasileiras, públicas federais e privadas, no período de 2019 a 2023, considerando indicadores de acesso, permanência, conclusão e qualidade educacional. Adicionalmente, analisou-se a influência de diferentes estratégias de financiamento público sobre o desempenho institucional, com destaque para o papel do Fundo de Financiamento Estudantil (Fies) como política de apoio ao acesso ao ensino superior privado. A investigação foi conduzida por meio da Análise Envoltória de Dados (DEA) em múltiplas etapas, sob retornos variáveis de escala e orientação a produtos, complementada por procedimentos de correção por bootstrap e pela aplicação do Índice de Produtividade de Malmquist, permitindo distinguir efeitos associados a melhorias de eficiência gerencial e a avanços tecnológicos. Os escores de eficiência foram posteriormente relacionados a variáveis ambientais e institucionais — como localização geográfica, tempo de existência, PIB regional, densidade demográfica e participação no Fies — por meio de técnicas de aprendizado de máquina supervisionado. A amostra compreendeu dados do período de 2019 a 2023, provenientes do Censo da Educação Superior, do Índice Geral de Cursos (IGC) e de indicadores socioeconômicos do IBGE. Os resultados indicaram que as IES públicas apresentaram maior estabilidade nos níveis de eficiência, enquanto as privadas exibiram maior heterogeneidade de desempenho. Observou-se ainda que fatores regionais e estruturais influenciam a eficiência relativa das instituições e que a participação no Fies está associada a padrões distintos de desempenho entre as IES privadas. Conclui-se que, embora o financiamento estudantil amplie o acesso ao ensino superior, sua atuação não assegura, por si só, maior eficiência técnica das instituições beneficiadas.

Palavras-chave: ensino superior; eficiência técnica; DEA em rede; Fies; políticas públicas.

ABSTRACT

MELLO, Daniel da Silva. *EFFICIENCY OF HIGHER EDUCATION INSTITUTIONS IN BRAZIL: A DEA Analysis on Resource Allocation*. 2026. 205f. Dissertation (Master's degree in Accounting Sciences) – University of Brasília, Brasília, 2026.

This study aimed to evaluate the technical efficiency applied to the production process of Brazilian Higher Education Institutions (HEIs), both federal public and private, during the period from 2019 to 2023, considering indicators of access, retention, completion, and educational quality. Additionally, the study examined the influence of different public financing strategies on institutional performance, with particular emphasis on the role of the Student Financing Fund (Fies) as a policy instrument supporting access to private higher education. The investigation was conducted using Data Envelopment Analysis (DEA) in multiple stages, assuming variable returns to scale and output orientation. The model was complemented by bootstrap correction procedures and the application of the Malmquist Productivity Index, enabling the identification of effects associated with managerial efficiency improvements and technological progress. Efficiency scores were subsequently related to environmental and institutional variables—such as geographic location, institutional age, regional GDP, population density, and participation in Fies — through supervised machine learning techniques. The sample comprised data from 2019 to 2023 obtained from the Higher Education Census, the General Course Index (IGC), and socioeconomic indicators from the Brazilian Institute of Geography and Statistics (IBGE). The results indicate that public HEIs show greater stability in efficiency levels, whereas private institutions display greater heterogeneity in performance. It was also observed that regional and structural factors influence the relative efficiency of institutions and that participation in Fies is associated with distinct performance patterns among private HEIs. The study concludes that, although student financing policies expand access to higher education, they do not, by themselves, ensure higher technical efficiency among the beneficiary institutions.

Keywords: higher education; technical efficiency; network DEA; student loan program (Fies); public policy.

LISTA DE FIGURAS

Figura 1 - Índice de Produtividade em diferentes períodos	27
Figura 2 - Curvas CCR (CRS) e BCC (VRS) e decomposição eficiência técnica	31
Figura 3 - Medida de Eficiência Técnica orientada aos insumos	31
Figura 4 - Eficiência Técnica orientada aos produtos	32
Figura 5 - Realização do cálculo DEA em três etapas (Network DEA).....	61
Figura 6 - Distribuição das IES do Brasil (2019 e 2023).....	77

LISTA DE GRÁFICOS

Gráfico 1 - Boxplots das Variáveis Originais	82
Gráfico 2 - Boxplots das Variáveis Tratadas	82
Gráfico 3 - Boxplots das Eficiências por Etapas (CRS)	93
Gráfico 4 - Boxplots das Eficiências por Etapas (VRS).....	93
Gráfico 5 - Boxplots Eficiências corrigidas nas três etapas (2023)	100
Gráfico 6 - Dispersão de eficiências com intervalo de confiança (Etapa 1 – 2023)	101
Gráfico 7 - Dispersão de eficiências com intervalo de confiança (Etapa 2 – 2023)	101
Gráfico 8 - Dispersão de eficiências com intervalo de confiança (Etapa 3 – 2023)	102
Gráfico 9 - Médias anuais do IM e componentes por etapas, 2019–2023	117

LISTA DE QUADROS

Quadro 1 - Síntese da literatura empírica sobre eficiência em IES	42
Quadro 2 - Variáveis utilizadas em Estudos Anteriores	50
Quadro 3 - Variáveis escolhidas – Descrição e Metodologia	51
Quadro 4 - Variáveis Ambientais.....	69

LISTA DE TABELAS

Tabela 1 - Caracterização das IES da amostra (2019–2023).....	75
Tabela 2 - Análise descritiva dos insumos e produtos – originais	79
Tabela 3 - Análise descritiva dos insumos e produtos – tratados.....	80
Tabela 4 - IES supereficientes Etapa 1 (DEA-VRS OO, 2019–2023)	84
Tabela 5 - IES supereficientes Etapa 2 (DEA-VRS OO, 2019–2023)	86
Tabela 6 - IES supereficientes Etapa 3 (DEA-VRS OO, 2019–2023)	90
Tabela 7 - Resultados do teste KS para os anos de 2019 a 2023	92
Tabela 8 - Resultados do teste KW para os anos de 2019 a 2023	94
Tabela 9 - Ranking de Eficiências NDEA Clássico (2023)	97
Tabela 10 - Ranking de Eficiências NDEA Bootstrap (2023).....	99
Tabela 11 - Gastos Direto e Indireto, 2019–2023.....	103
Tabela 12 - Estatísticas descritivas das eficiências por rede e ano (2019–2023) ...	103
Tabela 13 - Resultados dos modelos de aprendizado de máquina	105
Tabela 14 - Importância das variáveis ambientais – XGBoost	108
Tabela 15 - Resultados da Regressão Beta para a Eficiência Global (EG).....	111

LISTA DE ABREVIATURAS E SIGLAS

ACG	Avaliação de Cursos de Graduação
BCC	BANKER, CHARNES e COOPER (modelo DEA com retornos variáveis de escala – VRS)
CCR	CHARNES, COOPER e RHODES (modelo DEA com retornos constantes de escala – CRS)
CPC	Conceito Preliminar de Curso (Sinaes)
CRS	Constant Returns to Scale (retornos constantes de escala)
DEA	Data Envelopment Analysis (Análise Envoltória de Dados)
DMU	Decision Making Unit (Unidade Tomadora de Decisão)
EAD	Educação a Distância
EC	Emenda Constitucional
EE	Eficiência de Escala
ETP	Eficiência Técnica Pura
Enade	Exame Nacional de Desempenho dos Estudantes
Fies	Fundo de Financiamento Estudantil
IBGE	Instituto Brasileiro de Geografia e Estatística
IES	Instituições de Ensino Superior
IFES	Instituições Federais de Ensino Superior
IGC	Índice Geral de Cursos
IM	Índice de Produtividade de Malmquist
Inep	Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira
MEC	Ministério da Educação
NDEA	Network Data Envelopment Analysis (DEA em Rede)
PECH	Componente de Eficiência Técnica Pura do IM
SECH	Componente de Eficiência de Escala do IM
Sinaes	Sistema Nacional de Avaliação da Educação Superior
TC	Technical Change (Mudança Tecnológica, componente do IM)
UnB	Universidade de Brasília
VRS	Variable Returns to Scale (retornos variáveis de escala)

SUMÁRIO

1	INTRODUÇÃO	15
1.1	OBJETIVOS.....	19
1.2	JUSTIFICATIVA	21
2	REVISÃO DA LITERATURA (TEÓRICA E EMPÍRICA)	24
2.1	REFERENCIAL TEÓRICO: CONCEITOS E MEDIDAS DA EFICÁCIA, PRODUTIVIDADE E EFICIÊNCIA.....	24
2.1.1	Eficácia	24
2.1.2	Produtividade	26
2.1.3	Eficiência	28
2.2	DEA NO CÁLCULO DA EFICIÊNCIA	33
2.3	CONTEXTUALIZAÇÃO E CARACTERIZAÇÃO DO OBJETO DE ESTUDO..	35
2.3.1	O Ensino Superior no Brasil: expansão e avaliação institucional	35
2.3.2	O Financiamento do Ensino Superior no Brasil	36
2.4	ESTUDOS ANTERIORES	39
2.5	HIPÓTESES DE PESQUISA	43
2.5.1	Diferenças de eficiência entre modelos institucionais	44
2.5.2	Efeitos do financiamento estudantil sobre a eficiência institucional	44
2.5.3	Influência de fatores institucionais e regionais na eficiência	45
2.5.4	Dinâmica temporal da produtividade institucional	45
3	MÉTODO DE PESQUISA	47
3.1	SELEÇÃO DE VARIÁVEIS	49
3.2	IDENTIFICAÇÃO DE OUTLIERS E SUPEREFICIÊNCIA.....	53
3.3	DEA-CCR E DEA-BCC	56
3.3.1	Network DEA	59
3.4	TESTE DE MODELOS	61
3.5	ANÁLISE DE GRUPOS	63
3.6	BOOTSTRAP.....	64
3.7	ANÁLISE DO IMPACTO DAS VARIÁVEIS AMBIENTAIS	65
3.8	ANÁLISE EXPLICATIVA DOS EFEITOS DO FIES POR BETA REGRESSION.....	71
3.9	AVALIAÇÃO DINÂMICA E ÍNDICE DE PRODUTIVIDADE DE MALMQUIST	72
4	ANÁLISE DOS RESULTADOS	74

4.1	ANÁLISE DESCRITIVA	74
4.2	IDENTIFICAÇÃO DE OUTLIERS	81
4.2.1	Supereficiência	83
4.3	TESTE DE MODELO (CRS OU VRS)	91
4.4	ANÁLISE DE GRUPOS (IES PÚBLICAS x PRIVADAS).....	93
4.5	NDEA-CLÁSSICO E BOOTSTRAP ORIENTADO AOS PRODUTOS	95
4.5.1	Cálculo da Eficiência a partir da utilização do bootstrap.....	98
4.5.2	Resultados de Eficiência entre Redes de Ensino Superior	102
4.6	FATORES EXÓGENOS	105
4.7	EFEITOS DO FIES SOBRE A EFICIÊNCIA DAS IES PRIVADAS	110
4.8	AVALIAÇÃO DINÂMICA E ÍNDICE DE PRODUTIVIDADE DE MALMQUIST	115
5	CONSIDERAÇÕES FINAIS	120
	REFERÊNCIAS	123
	APÊNDICE A – DMUS E VARIÁVEIS UTILIZADAS	139
	APÊNDICE B – RESULTADOS DE EFICIÊNCIA PARA AS TRÊS ETAPAS	141
	APÊNDICE C – RESULTADOS DE EFICIÊNCIA PARA AS TRÊS ETAPAS COM BOOTSTRAP	143
	APÊNDICE D – RESULTADOS DO IM PARA AS TRÊS ETAPAS.....	145
	APÊNDICE E – <i>SCRIPT</i> UTILIZADO NO SOFTWARE R-STUDIO.....	155
	APÊNDICE F – ESTUDOS EMPÍRICOS SOBRE EFICIÊNCIA EM IES.....	204

1 INTRODUÇÃO

As Instituições de Ensino Superior (IES) desempenham papel central no desenvolvimento socioeconômico e cultural, seja por sua função formadora de profissionais, seja por sua contribuição na pesquisa e extensão universitária (Oliveira; Morais, 2025).

As primeiras escolas superiores surgiram em 1808, com a chegada da família real portuguesa, quando foram criadas as Escolas de Cirurgia e Anatomia em Salvador e no Rio de Janeiro, além da Academia da Guarda Marinha. Poucos anos depois, seguiram-se a Academia Real Militar em 1810, o curso de Agricultura em 1814 e a Real Academia de Pintura e Escultura em 1816 (Cunha, 2007). Durante o Império, o ensino superior permaneceu restrito às elites.

As tentativas de criação de universidades no Brasil, desde o período colonial até o final do século XIX, não obtiveram êxito, refletindo resistências tanto da Coroa portuguesa quanto de setores da própria sociedade (Fávero, 2006). Nesse contexto, até a Proclamação da República em 1889 prevaleceu o modelo de faculdades isoladas, voltadas à formação de profissionais liberais, com prestígio social e acesso a cargos públicos (Martins, 2002). Ainda no século XIX, instituições científicas como o Museu Nacional e o Observatório Nacional consolidaram-se como centros de produção científica, embora a expansão do ensino superior fosse limitada pela escassez de recursos estatais (Oliveira; Morais, 2025).

No início do século XX, o ensino superior ampliou-se de 24 estabelecimentos em 1900 para 133 em 1930, com forte crescimento na década de 1920 (García, 1998). O marco inicial do modelo universitário foi a criação da Universidade do Paraná (1912) e da Universidade do Rio de Janeiro (1920), que integravam escolas isoladas (Cunha, 2007). A Reforma Francisco Campos de 1931 introduziu medidas de modernização, como seriação, frequência obrigatória e sistema de avaliações, mas manteve a possibilidade de cobrança de anuidades, reforçando o caráter seletivo do ensino superior (Dallabrida, 2009).

A criação da Universidade de São Paulo (1934) é considerada um marco por reunir intelectuais estrangeiros e estruturar-se no tripé ensino–pesquisa–extensão, configurando-se como um centro de referência acadêmica e científica no país (Paula, 2009). Entre 1931 e 1945, disputas entre grupos laicos e católicos resultaram na criação de universidades confessionais.

Entre 1940 e 1960, o número de matrículas no ensino superior triplicou, alcançando mais de 226 mil estudantes em 1960, distribuídos entre instituições públicas e privadas, em consonância com a expansão populacional e o processo de industrialização (Martins, 2002). Nesse contexto, a Lei de Diretrizes e Bases da Educação Nacional (LDB) de 1961 fixou princípios gerais sem impor a obrigatoriedade do modelo universitário (Cardoso; Sampaio, 1994).

O regime militar instaurado em 1964 impactou diretamente o ensino superior brasileiro, impondo forte vigilância sobre as universidades públicas e, ao mesmo tempo, ampliando o espaço do setor privado. Nesse contexto, o Ato Institucional n. 5, de 1968, intensificou a repressão à comunidade acadêmica (Martins Filho, 1987).

Ainda em 1968, a Reforma Universitária promoveu mudanças estruturais significativas: extinguiu as cátedras vitalícias, instituiu departamentos, reorganizou o ensino por créditos e ciclos e regulamentou a pós-graduação *stricto sensu*, com base no Parecer CFE n. 977/65, de Newton Sucupira (Martins, 2009). Embora esses avanços tenham modernizado o sistema universitário, a reforma também reforçou os mecanismos de controle do Estado sobre as universidades, reafirmando a lógica centralizadora do período (Martins Filho, 1987).

A Constituição Federal de 1988 fortaleceu a educação como direito de todos e dever do Estado, o que promoveu a ampliação das universidades públicas, embora impondo novos desafios à gestão eficiente dos recursos públicos (Pereira, 2023).

O crescimento econômico das décadas de 1960 e 1970 elevou significativamente a demanda por profissionais qualificados, o que resultou na expansão das matrículas no ensino superior de 93.902 em 1960 para 1,34 milhão em 1980 (García, 1998). Nesse período, o setor privado ampliou de forma expressiva sua participação, passando de 50,5% das matrículas em 1970 para 63,3% em 1980 (Sampaio, 2000).

Com o processo de abertura política no final dos anos 1970 e a redemocratização na década seguinte, novas dinâmicas se estabeleceram no sistema de ensino superior. Durante as décadas de 1980 e 1990, o ensino noturno consolidou-se como um importante mecanismo de acesso para estudantes trabalhadores, ainda que marcado por condições precárias de permanência.

Paralelamente, o setor privado reforçou seu protagonismo, concentrando a maior parte das matrículas e consolidando-se como o principal responsável pela expansão da oferta de vagas (Oliveira; Bittar; Lemos, 2010).

No contexto das reformas educacionais e da modernização do Estado brasileiro, o governo Fernando Henrique Cardoso (1995–2002) marcou uma etapa decisiva na consolidação das políticas de avaliação e financiamento do ensino superior. Nesse período, ampliaram-se os instrumentos de regulação e monitoramento do sistema educacional.

Entre as principais iniciativas destacam-se a criação do Exame Nacional do Ensino Médio (Enem), a estruturação dos sistemas nacionais de avaliação e a instituição do Fundo de Financiamento Estudantil (Fies) em 1999, regulamentado posteriormente pela Lei n. 10.260, de 2001, concebido como instrumento de ampliação do acesso ao ensino superior privado (Pontuschka, 2016; Souza, 2021).

Ainda que, em 1994, o setor privado já concentrasse 69% das matrículas, o ensino público manteve a liderança na produção científica e na qualidade da formação (Sampaio, 2000). Entretanto, as transformações ocorridas na década de 1990 consolidaram uma lógica de expansão privatista, amparada por políticas de financiamento estudantil e incentivos fiscais, que passaram a constituir a principal estratégia de democratização do acesso ao ensino superior (Souza, 2021).

A análise da eficiência do ensino superior brasileiro exige a consideração da dimensão fiscal associada às políticas públicas de financiamento educacional. Nesse contexto, o Fies, criado em 1999 e ampliado a partir da década de 2010, constitui um dos instrumentos utilizados pelo Estado para ampliar o acesso ao ensino superior privado, mobilizando volumes expressivos de recursos públicos.

A relevância fiscal e institucional dessa política torna pertinente examiná-la como um dos fatores que podem influenciar o desempenho das Instituições de Ensino Superior no sistema educacional brasileiro. Estimativas indicam que, entre 2010 e 2023, os desembolsos ultrapassaram dezenas de bilhões de reais, configurando o Fies como uma das principais estratégias de expansão do setor privado no ensino superior brasileiro contemporâneo (Becker; Mendonça, 2021; Souza, 2021; Pontuschka, 2016).

No caso das Instituições de Ensino Superior (IES) públicas, os dispêndios governamentais representam a principal fonte de manutenção e expansão do sistema. Dados do Painel Estatístico do Censo da Educação Superior (Brasil, [s.d.]c) indicam que, entre 2019 e 2023, os gastos públicos médios anuais federais somaram mais de R\$ 17 bilhões, evidenciando o peso orçamentário da oferta pública direta no ensino superior brasileiro.

Tanto o financiamento indireto quanto o investimento direto nas universidades públicas federais absorvem volumes significativos de recursos públicos, ainda que apresentem trajetórias distintas. Após o auge de desembolsos verificado em meados da década de 2010, o Fies passou a registrar redução progressiva nos repasses federais, reflexo de ajustes fiscais e de maior seletividade na concessão de financiamentos (Brasil, [s.d.]b; [s.d.]c).

Em contrapartida, os investimentos nas Instituições de Ensino Superior públicas mantiveram-se em patamar relativamente mais estável, com tendência de crescimento gradual ao longo do período analisado, reforçando o papel do Estado como principal mantenedor do sistema público de ensino superior (Brasil, [s.d.]b; [s.d.]c).

Diversos estudos têm investigado a eficiência das Instituições de Ensino Superior por meio da Análise Envoltória de Dados (DEA) e de técnicas complementares, evidenciando diferenças relevantes de desempenho entre instituições e sistemas educacionais. Pesquisas internacionais apontam que fatores institucionais, organizacionais e regionais influenciam significativamente os níveis de eficiência universitária (Johnes, 2006; Parteka; Wolszczak-Derlacz, 2013; Agasisti et al., 2021). No contexto brasileiro, evidências empíricas também indicam variações de eficiência entre universidades públicas federais, associadas a diferenças de escala, perfil acadêmico e estrutura institucional (Costa et al., 2012; Santos; Silva Júnior; Nunes, 2022).

Paralelamente, a literatura sobre financiamento do ensino superior brasileiro tem examinado os efeitos do Fies sobre o acesso e a sustentabilidade das instituições privadas. Esses estudos mostram que o programa contribuiu para ampliar o acesso ao ensino superior, mas também gerou desafios relacionados ao custo fiscal e à sustentabilidade do modelo de financiamento (Pontuschka, 2016; Souza, 2021; Becker; Mendonça, 2021).

Além disso, a literatura aponta que investimentos adicionais em instituições ineficientes não necessariamente elevam seu desempenho e podem, ao contrário, resultar em desperdício de recursos públicos. Nesse sentido, Rosano-Peña, Albuquerque e Carvalho (2012) destacam a importância de identificar previamente as fontes de ineficiência institucional antes da ampliação de aportes financeiros, de modo a assegurar uma alocação mais eficiente dos recursos públicos.

Apesar dos avanços empíricos observados na literatura, ainda são escassos os estudos que analisam de forma integrada a eficiência relativa entre o financiamento público direto das universidades federais e o financiamento estudantil indireto via Fies. Também são limitadas as investigações que combinam abordagens estáticas e dinâmicas de eficiência, incorporam variáveis ambientais e examinam o papel das políticas públicas de financiamento na explicação das diferenças de desempenho institucional ao longo do tempo. Essa lacuna limita a capacidade de gestores públicos e formuladores de políticas de compreender de que maneira diferentes instrumentos de financiamento são convertidos em resultados educacionais no sistema de ensino superior brasileiro.

Diante desse contexto, estabelece-se o problema central desta pesquisa, formulado na seguinte pergunta:

Qual é a diferença de eficiência relativa entre as Instituições de Ensino Superior públicas e privadas no Brasil, considerando distintos mecanismos de financiamento público — investimento direto nas universidades federais e financiamento estudantil indireto por meio do Fies?

A partir dessa questão, busca-se analisar de que forma diferentes estratégias de financiamento público estão associadas ao desempenho institucional das IES, bem como discutir suas implicações para a alocação de recursos públicos e para o aprimoramento das políticas educacionais voltadas à expansão, equidade e sustentabilidade do ensino superior brasileiro.

1.1 OBJETIVOS

O objetivo principal deste trabalho é avaliar a eficiência técnica aplicada ao processo produtivo das Instituições de Ensino Superior (IES) brasileiras, públicas e privadas, no período de 2019 a 2023, considerando indicadores de acesso, permanência, conclusão e qualidade educacional. Neste estudo, o termo “processo produtivo” é utilizado em sentido analítico, conforme a literatura de eficiência organizacional, referindo-se à capacidade das instituições de transformar insumos educacionais — como recursos financeiros, infraestrutura e corpo docente — em resultados acadêmicos e formativos.

No intuito de cumprir esse objetivo geral, os seguintes objetivos específicos são estipulados de modo a se direcionar o trabalho proposto:

- Revisar a literatura, mapeando variáveis e metodologias aplicadas em estudos sobre eficiência e produtividade no ensino superior, destacando análises sobre o Fies, IES públicas e indicadores de qualidade institucional (como IGC e CPC);
- Mensurar a eficiência do processo produtivo das IES com identificação de outliers, retornos de escala e viés de mensuração;
- Examinar a influência de fatores ambientais e institucionais sobre a eficiência técnica das Instituições de Ensino Superior, considerando características regionais, estruturais e organizacionais;
- Avaliar o impacto da participação no Fies sobre a eficiência técnica das Instituições de Ensino Superior privadas, em comparação com as Instituições de Ensino Superior públicas;
- Analisar a evolução temporal da produtividade das Instituições de Ensino Superior brasileiras, distinguindo mudanças associadas a avanços tecnológicos e a melhorias de eficiência ao longo do período analisado.

Os objetivos delineados buscam preencher lacunas ainda presentes na literatura sobre eficiência e produtividade no ensino superior brasileiro, contribuindo não apenas para o avanço acadêmico, mas também para o aprimoramento das políticas públicas. Ao comparar a eficiência dos recursos aplicados diretamente nas IES públicas com aqueles destinados ao Fies, esta pesquisa pretende oferecer evidências empíricas que subsidiem gestores educacionais e formuladores de políticas na busca por uma alocação mais eficiente, equitativa e sustentável dos recursos públicos.

Sob a ótica da Administração Pública, adota-se uma visão sistêmica das Instituições de Ensino Superior como organizações públicas e privadas inseridas em um arranjo institucional complexo, no qual decisões de financiamento, regulação e gestão produzem efeitos sociais e econômicos relevantes. A análise dialoga com diferentes modelos de gestão pública — burocrático, gerencial e de governança — ao examinar como distintas estruturas institucionais e mecanismos de financiamento influenciam a capacidade das organizações educacionais de transformar recursos disponíveis em resultados educacionais.

Nesse contexto, a eficiência técnica aplicada ao processo operacional é tratada como um indicador de desempenho institucional associado às funções de

planejamento, orçamento e controle do setor público. Essa abordagem permite avaliar a qualidade da ação estatal à luz dos princípios da economicidade, eficiência e efetividade das políticas públicas voltadas ao ensino superior.

Sob a perspectiva da Contabilidade Aplicada ao Setor Público, a pesquisa contribui para a análise da eficiência do gasto público em educação superior ao deslocar o foco do controle estritamente financeiro para a avaliação de resultados e desempenho institucional. Ao examinar um instrumento de financiamento indireto que impacta o orçamento público sem se materializar diretamente como despesa orçamentária das Instituições Federais de Ensino, amplia-se o debate sobre a mensuração da eficiência na alocação de recursos públicos no setor educacional.

Nesse sentido, o estudo insere-se no campo da controladoria pública ao fornecer evidências empíricas que subsidiam o monitoramento, a avaliação e o aprimoramento de políticas públicas educacionais. A análise é conduzida à luz dos princípios da economicidade, eficiência e efetividade, contribuindo para a compreensão de como instrumentos de financiamento podem influenciar o desempenho institucional no ensino superior.

1.2 JUSTIFICATIVA

Desde o início dos anos 2000, as políticas públicas de expansão do ensino superior no Brasil passaram a combinar estratégias distintas de financiamento: de um lado, o investimento direto nas Instituições de Ensino Superior (IES) públicas, tradicionalmente responsáveis pela produção científica e pela formação acadêmica de qualidade; de outro, a ampliação do setor privado por meio do Fies, que mobilizou bilhões de reais em recursos públicos destinados ao custeio de matrículas privadas (Souza, 2021; Pontuschka, 2016).

No período recente, a literatura também aponta mudanças relevantes na composição do financiamento das Instituições Federais de Ensino Superior (IFES). Análises do orçamento federal indicam que, após um período de expansão observado até meados da década de 2010, as despesas discricionárias das IFES — relacionadas principalmente ao custeio e aos investimentos institucionais — passaram a apresentar retração relativa no orçamento total das universidades. Em 2014, essas despesas representavam aproximadamente 14,7% do orçamento das instituições, enquanto em 2019 essa participação foi reduzida para cerca de 9,2%, percentual inferior ao observado no início da série analisada (Reis; Cardoso; Adamczyk, 2022, p. 9).

Embora ambas as estratégias tenham contribuído para ampliar o acesso ao ensino superior, persistem incertezas quanto à eficiência relativa de cada modelo de financiamento, especialmente quando se consideram os custos fiscais, os resultados acadêmicos e os impactos sociais decorrentes dessas políticas (Becker; Mendonça, 2021). Essa questão torna-se particularmente relevante em contextos de restrição orçamentária, nos quais a alocação eficiente de recursos públicos passa a assumir papel central na formulação e na avaliação de políticas educacionais.

Do ponto de vista regulatório, a relevância desta pesquisa decorre das demandas de avaliação e fiscalização conduzidas pelo Ministério da Educação (MEC), particularmente no âmbito do Sistema Nacional de Avaliação da Educação Superior (Sinaes). A busca por transparência, qualidade e governança reforça a importância de compreender não apenas o desempenho das IES públicas, mas também os efeitos das políticas de financiamento estudantil, que sustentam grande parte das matrículas em instituições privadas (Brasil, [s.d.]).

No campo acadêmico, a contribuição deste estudo reside na integração de abordagens quantitativas, em especial a Análise Envoltória de Dados (DEA) — amplamente utilizada em estudos de eficiência (Farrell, 1957; Charnes; Cooper; Rhodes, 1978) — com técnicas de aprendizado de máquina (ML) e o Índice de Produtividade de Malmquist. Essa abordagem possibilita uma análise comparativa e multietápica do desempenho das IES, avançando em relação à literatura existente, que tradicionalmente examina a eficiência institucional de forma isolada, sem confrontar diretamente os modelos de financiamento público direto e indireto (Mota Júnior; Torres, 2016).

Em complemento à contribuição científica, o estudo apresenta também implicações práticas relevantes para a avaliação de políticas públicas e para a gestão do sistema educacional. A abordagem metodológica adotada permite:

- a) Incorporar múltiplos insumos e resultados (desejáveis e não desejáveis), ponderados conforme o perfil institucional, permitindo que a DEA estime endogenamente os pesos relativos de inputs e outputs;
- b) Testar a hipótese sobre o tipo de retornos de escala (CCR ou BCC), conforme Simar e Wilson (2002), mitigando distorções decorrentes da escolha arbitrária de tecnologia;

- c) Considerar a cadeia produtiva do ensino superior de forma integrada, identificando ineficiências específicas em etapas distintas — acesso, permanência e conclusão;
- d) Comparar as IES com benchmarks de melhores práticas, assegurando a comparabilidade entre instituições de diferentes portes e condições operacionais;
- e) Classificar cada instituição como eficiente ou ineficiente a partir de escores sintéticos, indicando aquelas que podem servir de referência para as demais;
- f) Basear-se em observações individuais, e não em médias agregadas, permitindo estabelecer metas de melhoria e trajetórias ótimas de desempenho;
- g) Empregar técnicas de aprendizado de máquina (ML) para analisar a influência de variáveis exógenas sobre os escores de eficiência, gerando estimativas ajustadas por fatores contextuais;
- h) Adotar uma perspectiva longitudinal por meio do Índice de Malmquist, possibilitando mensurar as mudanças de eficiência e produtividade no período de 2019 a 2023.

Dessa forma, a análise proposta nesta pesquisa busca contribuir para o aprimoramento do debate sobre a eficiência do gasto público no ensino superior brasileiro, oferecendo evidências empíricas que possam subsidiar gestores educacionais e formuladores de políticas públicas na busca por um equilíbrio mais eficiente entre o financiamento indireto via Fies e o investimento direto nas Instituições de Ensino Superior públicas.

Por fim, este trabalho está estruturado em cinco capítulos. Após esta introdução, o capítulo seguinte apresenta o referencial teórico; o terceiro descreve os procedimentos metodológicos adotados; o quarto apresenta e discute os resultados empíricos da pesquisa; e o quinto capítulo reúne as considerações finais.

2 REVISÃO DA LITERATURA (TEÓRICA E EMPÍRICA)

O presente estudo fundamenta-se na Teoria da Eficiência, que oferece um arcabouço conceitual sólido para analisar o desempenho relativo das Instituições de Ensino Superior (IES) e das políticas públicas voltadas à sua expansão. Essa teoria permite avaliar simultaneamente a relação entre recursos empregados e resultados alcançados, sendo particularmente adequada para comparar diferentes modelos de intervenção do Estado no ensino superior, como o financiamento indireto via Fies e o investimento direto nas IES públicas.

2.1 REFERENCIAL TEÓRICO: CONCEITOS E MEDIDAS DA EFICÁCIA, PRODUTIVIDADE E EFICIÊNCIA

2.1.1 Eficácia

O conceito de eficácia está diretamente associado ao cumprimento dos objetivos, missão ou metas institucionais, ainda que não envolva a consideração dos recursos empregados no processo. Para Giménez, Prior e Thieme (2007), a eficácia constitui condição necessária — embora não suficiente — para a obtenção do máximo rendimento, na medida em que somente organizações capazes de alcançar seus propósitos podem aspirar à otimização do uso dos recursos. Assim, ser eficaz não implica, por si só, ser eficiente, mas constitui condição prévia indispensável para que a eficiência possa ser efetivamente alcançada.

Importa destacar, entretanto, que a eficácia difere conceitualmente da eficiência e da produtividade. Enquanto a eficácia se restringe ao alcance dos objetivos propostos, a eficiência e a produtividade envolvem a relação entre resultados obtidos e recursos empregados. A eficiência busca avaliar se a instituição utiliza seus insumos da melhor forma possível para gerar determinado nível de resultados, ao passo que a produtividade examina a intensidade dessa relação, mensurando o quanto se produz com os recursos disponíveis. Dessa forma, eficácia, eficiência e produtividade são dimensões distintas, porém complementares, da avaliação de desempenho institucional.

Na literatura sobre avaliação de desempenho educacional, o conceito de eficácia é frequentemente apresentado como uma dimensão distinta da eficiência e da produtividade. Belloni (2000) compreende a eficácia como um critério de

desempenho vinculado ao cumprimento de metas e objetivos institucionais previamente definidos. De forma semelhante, Rosano-Peña (2008) define a eficácia como uma medida normativa orientada ao alcance dos objetivos estabelecidos pelas organizações. Sob essa perspectiva, um sistema que não atinge ou seleciona de forma inadequada seus objetivos apresenta desempenho ineficaz, independentemente da relação custo-benefício obtida.

O debate também aponta para a necessidade de critérios e indicadores adequados de avaliação. Savoie e Morin (2001) ressaltam que a eficácia constitui o núcleo do bom desempenho organizacional, mas é frequentemente subestimada, uma vez que muitas instituições ainda recorrem a métricas limitadas para mensurar seus resultados. Boueri *et al.* (2015) reforçam esse entendimento ao conceituarem eficácia como a capacidade de atingir os objetivos propostos.

A mensuração da eficácia pode ser ilustrada de maneira prática. Rosano-Peña e Gomes (2018) sugerem que ela seja compreendida pela relação entre resultados obtidos e possíveis. No setor público, deve concentrar-se não apenas no cumprimento da meta, mas também na adequação qualitativa e quantitativa dos serviços prestados. Os autores exemplificam com o caso de um Tribunal de Justiça que estabeleceu como meta a redução de 15% no número de processos pendentes em determinado período e alcançou 21%. Nesse exemplo, o índice de eficácia foi de 1,4, indicando desempenho 40% superior à meta estabelecida. Por outro lado, índices inferiores a 1 indicam ineficácia; se uma instituição apresenta valor de 0,8, por exemplo, significa que deve melhorar seus resultados em 20% para ser considerada eficaz.

Entretanto, a avaliação torna-se mais complexa quando múltiplas metas são perseguidas simultaneamente, pois exige a ponderação de prioridades e a definição da importância relativa de cada objetivo. Nessas situações, diferentes perspectivas institucionais e juízos de valor podem influenciar a atribuição de pesos e critérios de avaliação, introduzindo elementos de arbitrariedade e subjetividade no processo de mensuração do desempenho (Belloni, 2000; Rosano-Peña, 2008).

Em síntese, a eficácia pode ser entendida como o grau de realização dos objetivos de uma organização. Trata-se de um requisito necessário — ainda que não suficiente — para a configuração da eficiência, uma vez que somente após o cumprimento dos objetivos é possível avaliar a adequação e a racionalidade no uso dos recursos empregados. Embora ambos os conceitos estejam relacionados, a satisfação plena de um não implica automaticamente o alcance do outro, reforçando

a importância de análises que considerem simultaneamente os resultados obtidos e os recursos utilizados no processo.

2.1.2 Produtividade

O conceito de produtividade ocupa papel central nos estudos de desempenho, sendo tradicionalmente definido como a razão entre produtos gerados e insumos empregados no processo de produção (Sharpe, 2002). Essa relação, expressa na forma de uma razão simples, mede quantos bens ou serviços podem ser produzidos a partir de determinada quantidade de recursos, sejam eles humanos, financeiros ou materiais.

Para Coelli *et al.* (2005), a produtividade é uma métrica de desempenho, de modo que valores mais elevados da razão produto/insumo indicam melhor aproveitamento dos recursos disponíveis. Essa lógica se aplica a diferentes contextos: no ensino superior, pode ser mensurada pelo número de alunos formados em relação ao corpo docente; na agricultura, pelo volume de grãos colhidos por hectare; e no setor público, pela quantidade de processos julgados por magistrado em determinado período.

A análise da produtividade pode assumir também uma dimensão temporal. Tupy e Yamaguchi (1998) ressaltam que a evolução da produção pode ser estudada comparando-se a produtividade em períodos distintos, o que permite verificar variações positivas ou negativas de desempenho ao longo do tempo. Esse tipo de análise é fundamental para compreender o impacto de inovações tecnológicas ou mudanças organizacionais.

Formalmente, a produtividade pode ser representada matematicamente como:

$$Produtividade = \frac{y}{x} \quad (1)$$

em que y representa a quantidade de produtos gerados e x corresponde aos insumos empregados.

Em contextos mais complexos, envolvendo múltiplos insumos e produtos, a mensuração parcial — isto é, a relação entre apenas um insumo e um produto — pode gerar conclusões distorcidas, já que desconsidera interações entre fatores. Nesse cenário, surge o conceito de Produtividade Total dos Fatores (PTF), que pondera todos os insumos e produtos simultaneamente, fornecendo uma visão mais abrangente do desempenho organizacional (Coelli *et al.*, 2005).

A formulação matemática da PTF pode ser representada da seguinte forma:

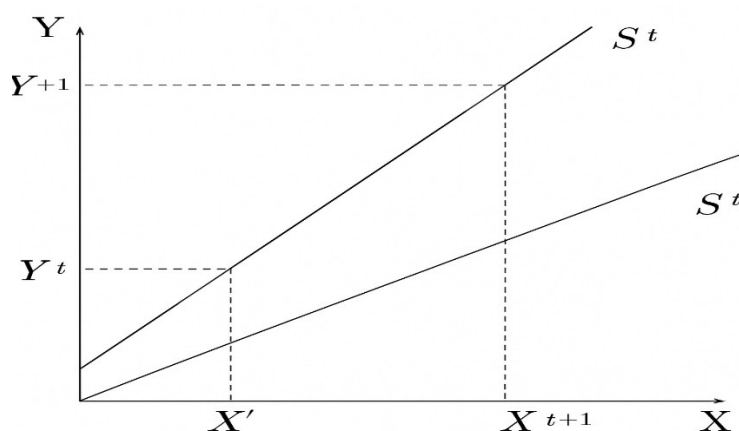
$$PTF = \frac{\sum_{s=1}^S u_s y_s}{\sum_{m=1}^M v_m x_m} \quad (2)$$

em que u_s e v_m são, respectivamente, os pesos ou preços atribuídos aos produtos (y) e insumos (x).

A análise da PTF permite identificar duas dimensões críticas: a) os avanços tecnológicos, conhecidos como *frontier-shift effect*, que ocorrem quando inovações deslocam a fronteira de produção para cima, permitindo maiores resultados com os mesmos recursos; e b) o *catch-up effect*, que mede a capacidade das unidades produtivas de se aproximarem das melhores práticas já existentes (Bogetoft; Otto, 2011).

Essas ideias podem ser representadas graficamente (Figura 1), que ilustra o crescimento da produtividade entre os períodos t e $t+1$ por meio do deslocamento da fronteira tecnológica, representada pelas curvas S^t e S^{t+1} . Observa-se que, para um mesmo nível de insumo, a produção possível no período $t+1$ é superior à do período anterior, evidenciando progresso tecnológico. Esse deslocamento da fronteira — e não uma mudança na posição relativa da unidade produtiva — indica que o aumento observado no produto (Y) decorre de avanços na tecnologia ou na organização dos processos produtivos, que ampliam a capacidade de transformar insumos em resultados. Assim, a figura demonstra que o crescimento da produtividade resulta fundamentalmente da mudança tecnológica, um dos componentes clássicos da decomposição do Índice de Malmquist (Färe *et al.*, 1994).

Figura 1 - Índice de Produtividade em diferentes períodos



Fonte: Adaptação de Vicente, 2004.

De modo aplicado, no campo da educação superior, a mensuração da produtividade permite avaliar em que medida as IES conseguem converter recursos humanos e financeiros em resultados tangíveis, como número de concluintes, produção científica ou inserção social (Johnes, 2006; Agasisti *et al.*, 2021).

Por fim, destaca-se que a análise da produtividade não se encerra em si mesma, mas constitui passo fundamental para a compreensão da eficiência. Isso porque, na perspectiva da gestão de processos, a maximização dos resultados e a minimização dos insumos representam princípios basilares da análise econômica e da avaliação de desempenho (Cooper; Seiford; Tone, 2007). Essa conexão prepara o terreno para a subseção seguinte, dedicada à discussão do conceito de eficiência.

2.1.3 Eficiência

A discussão sobre eficiência adquire maior consistência quando situada no marco da teoria da produção. A produção pode ser entendida como o processo de transformação de insumos (inputs) em produtos (outputs) a partir de determinada tecnologia. Esse raciocínio não se limita ao setor fabril ou industrial, mas abrange igualmente serviços, finanças e, no caso da educação superior, a combinação de recursos humanos, físicos e financeiros para gerar resultados em ensino, pesquisa e extensão (Mas-Colell; Whinston; Green, 1995).

A relação entre entradas e saídas pode ser formalmente expressa por uma função de produção:

$$y = f(x_1, x_2, \dots, x_n) \quad (3)$$

em que y representa a quantidade agregada de produtos e x_i são os insumos empregados. A partir dessa formulação, deriva-se o conceito de Conjunto de Possibilidades de Produção (CPP), definido como o conjunto de combinações viáveis de produtos e insumos em dado momento histórico, condicionado pela tecnologia disponível. As unidades de tomada de decisão (Decision Making Units – DMUs) que operam na fronteira desse conjunto são classificadas como eficientes, ao passo que aquelas situadas abaixo da fronteira são consideradas ineficientes, por utilizarem insumos em excesso ou apresentarem resultados inferiores ao potencial (Rosano-Peña; Gomes, 2018).

O CPP possui propriedades fundamentais, como: a) insumos e produtos são não negativos; b) é um conjunto fechado, contendo combinações máximas de

produção na fronteira; c) o processo produtivo é irreversível; d) existe limite máximo de produção para os insumos disponíveis; e) pode apresentar retornos constantes, crescentes ou decrescentes de escala; f) sua fronteira pode ter formato convexo, em consonância com a lei dos rendimentos decrescentes, e g) o CPP abrange as DMU que operam de maneira ineficiente, isto é, aquelas que utilizam mais recursos do que o necessário para um determinado nível de produção e/ou produzem menos do que outras DMU usando os mesmos insumos (Mas-Colell; Whinston; Green, 1995).

Com base nesses fundamentos, é possível avançar para a diferenciação entre eficácia, produtividade e eficiência. A eficácia refere-se ao cumprimento dos objetivos institucionais, independentemente dos recursos empregados (Giménez; Prior; Thieme, 2007; Belloni, 2000; Rosano-Peña, 2008). A produtividade mede a intensidade da relação entre produtos e insumos, isto é, quantos resultados são obtidos a partir dos recursos aplicados (Sharpe, 2002; Coelli *et al.*, 2005). Já a eficiência compara o desempenho efetivo de uma DMU com aquele obtido pelas melhores práticas observadas, sendo, portanto, a síntese desses conceitos (Farrell, 1957; Koopmans, 1951).

Farrell (1957) foi pioneiro ao propor a decomposição da eficiência em duas dimensões: eficiência técnica, relacionada à capacidade de obter a produção máxima a partir dos insumos disponíveis (independentemente de preços), e eficiência alocativa, que considera a combinação ótima de insumos em termos monetários, ou seja, o menor custo possível para determinado nível de produção. A combinação dessas duas dimensões resulta na chamada eficiência econômica total.

A eficiência técnica, por sua vez, pode ser mensurada sob duas orientações: a) ao insumo, que busca a minimização equiproporcional dos insumos mantendo constante o nível de produto; ou b) ao produto, que visa à maximização equiproporcional dos outputs, mantendo constantes os insumos. Essas formas ficaram conhecidas como medidas radiais de Farrell e podem ser vistas como casos particulares do conceito de Koopmans (1951). Além disso, Shephard (1970) apresentou a eficiência técnica como o inverso da medida de distância radial, consolidando a base matemática desse conceito para o caso de múltiplos insumos e produtos.

Lovell (1993) complementa essa abordagem ao indicar três formas de mensurar a eficiência técnica: a) comparando o produto observado com o produto

potencial máximo; b) comparando o insumo observado com o insumo mínimo necessário para determinado produto; ou c) por uma combinação dos dois critérios.

A eficiência técnica corresponde à capacidade de uma unidade produtiva gerar o máximo nível de output a partir de uma dada quantidade de inputs — ou, inversamente, utilizar a quantidade mínima de insumos para produzir um nível fixo de produto. Essa concepção foi originalmente formalizada por Michael J. Farrell, que estabeleceu as bases analíticas para a mensuração da eficiência produtiva. No contexto da Análise Envoltória de Dados (DEA), essa medida pode ser decomposta em duas dimensões: a eficiência técnica pura e a eficiência de escala (Charnes; Cooper; Rhodes, 1978; Banker; Charnes; Cooper, 1984).

A eficiência técnica pura, estimada sob a hipótese de retornos variáveis de escala (VRS), capta exclusivamente o desempenho gerencial e organizacional da unidade produtiva, isolando os efeitos relacionados ao porte. Já a eficiência de escala avalia se a unidade de decisão (DMU) opera em tamanho adequado, examinando em que medida desvios em relação à escala ótima — aquela que maximiza a produtividade total dos fatores — influenciam o nível global de eficiência (Banker; Charnes; Cooper, 1984).

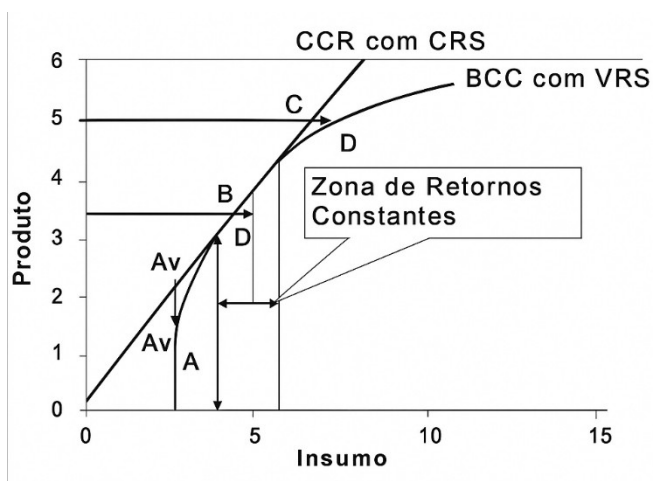
Nesse sentido, instituições subdimensionadas ou superdimensionadas podem apresentar eficiência técnica pura, mas ainda assim revelar ineficiência de escala, uma vez que operam fora do ponto em que retornos crescentes, constantes ou decrescentes de escala são economicamente desejáveis. A articulação entre essas duas componentes permite compreender se eventuais ineficiências decorrem de problemas gerenciais, de inadequação do porte ou de ambos (Cooper; Seiford; Tone, 2007).

Esses conceitos podem ser ilustrados graficamente, permitindo visualizar a separação entre as fontes de ineficiência e a posição das unidades produtivas em relação à fronteira e à escala ótima. A Figura 2 apresenta simultaneamente as fronteiras de eficiência sob retornos constantes de escala (CCR–CRS) e sob retornos variáveis de escala (BCC–VRS), permitindo visualizar a decomposição da eficiência técnica em eficiência técnica pura e eficiência de escala.

A curva BCC representa a fronteira eficiente sob VRS, refletindo exclusivamente o desempenho gerencial das unidades produtivas, enquanto a curva CCR representa a fronteira sob CRS, incorporando também os efeitos de escala (Charnes; Cooper; Rhodes, 1978; Banker; Charnes; Cooper, 1984). A diferença entre

as duas fronteiras para uma mesma DMU indica o componente de eficiência de escala. A zona de retornos constantes evidencia o intervalo em que as duas fronteiras coincidem.

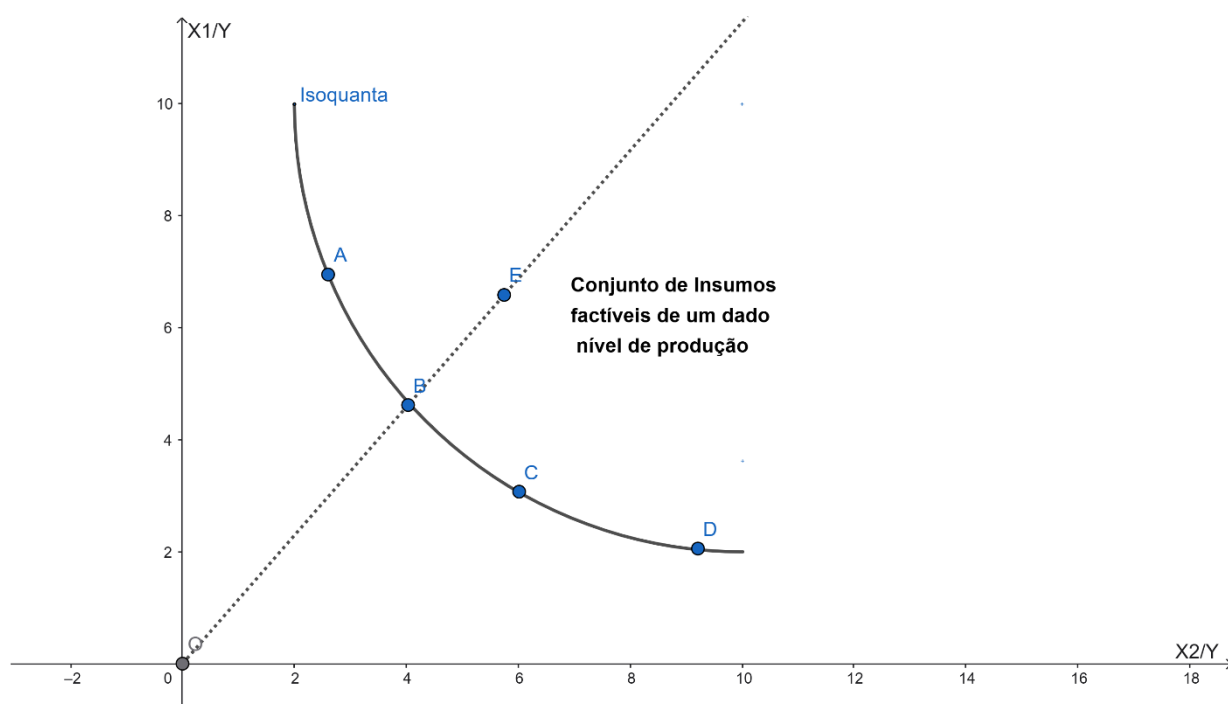
Figura 2 - Curvas CCR (CRS) e BCC (VRS) e decomposição eficiência técnica



Fonte: Adaptado de Rosano-Peña e Gomes, 2018.

Na figura 3, DMUs localizadas na fronteira do CPP são classificadas como eficientes, enquanto aquelas posicionadas acima da isoquanta são ineficientes, pois utilizam insumos em excesso. A distância radial entre a unidade observada e a fronteira representa a ineficiência técnica pura.

Figura 3 - Medida de Eficiência Técnica orientada aos insumos



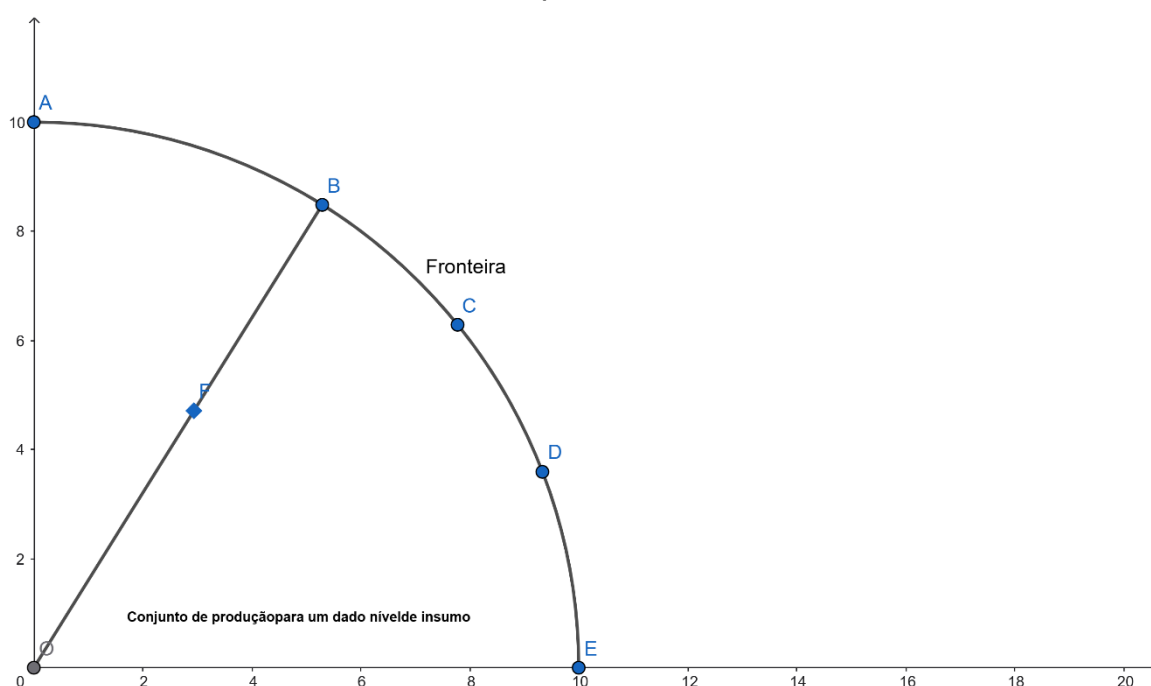
Fonte: Adaptado de Rosano-Peña e Gomes, 2018.

Na Figura 3, as DMUs eficientes em termos técnicos (A, B, C e D) formam a fronteira do CPP e as ineficientes colocam-se acima desta fronteira (E).

Se uma DMU utiliza quantidades de insumos definidas pelo ponto E para produzir uma unidade de produto y, a ineficiência técnica será representada pela distância \overline{BE} , que representa o menor montante proporcional que todos os inputs poderiam ser reduzidos sem alterar o nível de produção y.

A quantificação da ineficiência, conforme Farrell (1957), é normalmente expressa pela relação $\frac{\overline{OB}}{\overline{OE}}$.

Figura 4 - Eficiência Técnica orientada aos produtos



Fonte: Adaptado de Rosano-Peña e Gomes, 2018.

Na Figura 4, as DMUs tecnicamente eficientes (A, B, C, D e E) compõem a fronteira do Conjunto de Possibilidades de Produção (CPP), representando as melhores práticas observadas para diferentes níveis de insumo. Já a unidade F situa-se abaixo da fronteira, evidenciando o seu caráter ineficiente, pois produz um volume de output inferior ao potencial permitido pelos insumos empregados.

Tal ineficiência técnica orientada aos produtos será representada pela distância BF, que corresponde ao incremento proporcional mínimo que todos os outputs poderiam ser ampliados sem alterar o nível de insumo utilizado. O ponto B, localizado na fronteira eficiente, indica o nível máximo de produto que a DMU poderia alcançar mantendo constante a quantidade de insumos observada. Esse segmento expressa

o incremento proporcional de outputs que poderia ser obtido — mantendo constante o insumo utilizado — para que a unidade atinja o desempenho da fronteira.

Conforme Farrell (1957), a eficiência técnica orientada aos produtos pode ser quantificada pela razão $\frac{OF}{OB}$, em que: OF é o nível de output efetivamente produzido pela DMU no ponto F; e OB é o nível eficiente de output correspondente ao ponto B na fronteira.

Essa razão representa a proporção entre o produto observado e o produto máximo possível. Quanto mais distante F estiver da fronteira (isto é, quanto maior a distância BF), menor será o valor de $\frac{OF}{OB}$, indicando maior ineficiência técnica.

Do ponto de vista metodológico, a eficiência pode ser estimada por métodos paramétricos e não paramétricos. O primeiro, mais tradicional na economia, baseia-se em regressões que exigem a definição prévia de uma relação funcional entre insumos e produtos, fornecendo resultados probabilísticos. Já os métodos não paramétricos, como a Análise Envoltória de Dados (DEA), constroem a fronteira de eficiência por meio de programação matemática de otimização com restrições, sem necessidade de especificar funções prévias (Emrouznejad; Yang, 2018; Liu *et al.*, 2013).

Este estudo adota a DEA por sua flexibilidade em lidar com múltiplos insumos e outputs, bem como por permitir a comparação direta entre DMUs de diferentes portes e perfis. Além disso, a utilização de técnicas como o bootstrap possibilita lidar com incertezas inerentes aos processos produtivos e viabiliza inferências estatísticas mais robustas (Simar; Wilson, 2007).

Em síntese, a eficiência representa não apenas a capacidade de transformar recursos em resultados da melhor forma possível, mas também o alinhamento da operação das instituições às melhores práticas observadas. Essa noção é particularmente relevante para a análise proposta, que busca comparar a eficiência relativa das IES públicas e do Fies no ensino superior brasileiro.

2.2 DEA NO CÁLCULO DA EFICIÊNCIA

A Teoria da Eficiência foi inicialmente apresentada por Farrell (1957), ao propor a mensuração do desempenho produtivo a partir da relação entre insumos e produtos. Essa abordagem serviu de base para o modelo de Análise Envoltória de Dados (DEA), desenvolvido por Charnes, Cooper e Rhodes (1978), que se consolidou como

ferramenta para avaliar a eficiência relativa de diferentes unidades de decisão (Decision Making Units – DMUs).

No âmbito das Instituições de Ensino Superior (IES), a DEA possibilita avaliar o desempenho de IES públicas e privadas que transformam recursos financeiros, humanos e físicos em resultados acadêmicos e sociais. Segundo Nazarko e Šaparauskas (2014), a metodologia permite identificar instituições que operam na fronteira da eficiência, as quais podem servir de referência para aquelas com desempenho inferior.

Embora tradicionalmente aplicada a indicadores quantitativos, a literatura tem enfatizado a incorporação de variáveis que medem a qualidade educacional. No caso brasileiro, destacam-se os indicadores do Sinaes, como o Índice Geral de Cursos (IGC) e o Conceito Preliminar de Curso (CPC), que complementam a análise de eficiência ao incluir dimensões de qualidade institucional (Brasil, [s.d.]).

A expansão da rede privada a partir dos anos 2000 trouxe novos desafios, sobretudo quanto à sustentabilidade financeira e à garantia de qualidade (Mota Júnior e Torres, 2016). Nesse contexto, a análise comparativa da eficiência entre IES públicas e políticas de financiamento estudantil, como o Fies, torna-se fundamental para compreender a relação entre acesso, qualidade e uso racional dos recursos públicos (Souza, 2021; Pontuschka, 2016; Becker; Mendonça, 2021).

Além da avaliação das IES em termos de eficiência técnica e qualidade institucional, a literatura também chama atenção para a magnitude dos recursos públicos alocados no financiamento estudantil. O Fies, por exemplo, mobilizou bilhões de reais desde sua criação em 1999, com forte expansão a partir de 2010 (Pontuschka, 2016; Souza, 2021). Esse volume de gastos suscita questionamentos relevantes sobre a efetividade da política em comparação ao investimento direto nas IES públicas. Assim, torna-se pertinente discutir em que medida a escolha pelo financiamento estudantil — em vez do fortalecimento das instituições públicas — representa uma decisão eficiente do ponto de vista econômico e social. Essa reflexão conecta a teoria da eficiência ao debate sobre políticas educacionais, fornecendo subsídios para avaliar não apenas o desempenho institucional, mas também a racionalidade das opções de alocação de recursos públicos.

2.3 CONTEXTUALIZAÇÃO E CARACTERIZAÇÃO DO OBJETO DE ESTUDO

Para compreender de forma aprofundada a problemática da eficiência no ensino superior brasileiro, é necessário situar o objeto de estudo a partir de duas vertentes centrais: de um lado, as Instituições de Ensino Superior (IES) públicas, historicamente responsáveis pela produção científica e pela formação acadêmica de referência no país; de outro, o Fies, política pública que, por meio da destinação de recursos estatais, ampliou significativamente o acesso ao setor privado.

Nesse contexto, esta seção apresenta inicialmente um panorama do desenvolvimento do ensino superior no Brasil, evidenciando a expansão das IES públicas e privadas e a consolidação de mecanismos de avaliação institucional, em especial o Sinaes. Em seguida, examina-se a trajetória do Fies, desde sua criação até as reformulações mais recentes, que redefiniram seu escopo de financiamento e produziram impactos relevantes sobre o sistema educacional.

A caracterização dessas duas dimensões — a oferta pública direta e o financiamento indireto por meio do Fies — é fundamental para delimitar o campo empírico da pesquisa. Busca-se compreender não apenas os efeitos em termos de acesso, permanência e conclusão, mas também as diferenças estruturais entre os modelos de financiamento, seus custos fiscais e suas repercussões sobre a qualidade institucional.

Dessa forma, este capítulo fornece os elementos necessários para contextualizar a análise comparativa proposta, situando as IES públicas e o Fies como objetos centrais de investigação no debate sobre a eficiência do ensino superior brasileiro.

2.3.1 O Ensino Superior no Brasil: expansão e avaliação institucional

O ensino superior brasileiro passou por transformações significativas nas últimas décadas, combinando a ampliação do acesso com a consolidação de mecanismos de avaliação institucional. Historicamente, as IES públicas desempenharam papel central na produção científica e na formação de recursos humanos de alta qualificação. No entanto, a partir dos anos 1990 e, sobretudo, dos anos 2000, observou-se forte expansão do setor privado, apoiada em políticas de financiamento estudantil e de inclusão social (Sampaio, 2000; Souza, 2021).

Nesse processo, a criação do Sinaes, instituído pela Lei n. 10.861/2004, representou um marco regulatório ao estabelecer critérios padronizados para mensurar qualidade institucional. Entre seus instrumentos principais, destacam-se o Exame Nacional de Desempenho dos Estudantes (Enade), a Avaliação de Cursos de Graduação (ACG) e a Avaliação Institucional, cujos resultados subsidiam a formulação de indicadores como o Conceito Preliminar de Curso (CPC) e o Índice Geral de Cursos (IGC) (Brasil, [s.d.]).

O Índice Geral de Cursos (IGC), em particular, tornou-se uma referência na mensuração da qualidade educacional no ensino superior brasileiro, uma vez que integra informações sobre desempenho discente, corpo docente, infraestrutura e resultados da pós-graduação. Embora apresente limitações, sua utilização permite comparações entre instituições e subsidia processos regulatórios e de financiamento conduzidos pelo Ministério da Educação (Brasil, [s.d.]). Dessa forma, o indicador torna-se relevante para análises de desempenho no ensino superior, pois associa dimensões quantitativas — relacionadas à expansão do acesso — a aspectos qualitativos vinculados à formação acadêmica e à estrutura institucional das IES.

Apesar dos avanços, a literatura evidencia desafios persistentes. Entre eles, a heterogeneidade entre instituições, a elevada taxa de abandono e a desigualdade regional no acesso e na qualidade (Mota Júnior; Torres, 2016; Becker; Mendonça, 2021). Nesse contexto, torna-se essencial compreender não apenas o desempenho das IES públicas, tradicionalmente associadas à excelência acadêmica, mas também a trajetória das instituições privadas, que concentram a maior parte das matrículas e dependem, em grande medida, de políticas públicas de financiamento.

Assim, o presente estudo incorpora a análise do Sinaes e do IGC como variáveis de qualidade institucional, ao mesmo tempo em que considera a expansão das redes pública e privada como elementos centrais para a caracterização do objeto investigado. Essa abordagem permite delimitar as bases empíricas sobre as quais será avaliada a eficiência relativa das IES públicas e do Fies como modelo de financiamento do ensino superior.

2.3.2 O Financiamento do Ensino Superior no Brasil

A análise da eficiência no ensino superior brasileiro precisa ser articulada com os dispêndios públicos destinados a esse nível educacional. Esses recursos se estruturam em duas frentes: os gastos diretos, realizados pela União na manutenção

das IES federais, e os gastos indiretos, representados por programas de financiamento estudantil, em especial o Fies. Essa distinção é destacada por Bielschowsky e Amaral (2022), que ressaltam a importância de compreender as diferentes modalidades de financiamento e seus impactos sobre o sistema como um todo.

O Fies foi instituído em 1999, em substituição ao antigo Crédito Educativo (Creduc), e regulamentado pela Lei n. 10.260, de 2001, com o objetivo de oferecer condições diferenciadas de financiamento a estudantes de baixa renda matriculados em instituições privadas, por meio de contratos com agentes financeiros credenciados pelo governo federal (Pontuschka, 2016).

Durante sua primeira década, o programa apresentou crescimento gradual. Entretanto, a partir de 2010 passou por uma ampla reformulação, com redução das taxas de juros, ampliação dos prazos de carência e eliminação da exigência de fiador para estudantes de menor renda. Essas mudanças resultaram em uma forte expansão: entre 2010 e 2015 foram firmados mais de 1,1 milhão de contratos, mobilizando cerca de R\$ 14 bilhões em recursos públicos federais (Souza, 2021; Pontuschka, 2016).

Apesar de sua contribuição para a democratização do acesso, a rápida expansão trouxe desafios de sustentabilidade fiscal e questionamentos sobre a eficiência do programa. Becker e Mendonça (2021) apontam que o Fies aumentou as matrículas e reduziu taxas de evasão, mas também prolongou o tempo médio de conclusão dos cursos e ampliou o risco de inadimplência. Críticas adicionais recaem sobre a dependência do setor privado em relação a subsídios públicos e sobre os custos fiscais crescentes, especialmente após 2015, quando foram implementadas medidas de contenção e maior seletividade na concessão dos financiamentos (Souza, 2021).

Outro aspecto central é a comparação entre o custo por aluno em instituições públicas e privadas. Segundo Bielschowsky e Amaral (2022), há uma confusão recorrente entre gasto por aluno — que inclui despesas totais da instituição divididas pelo número de matrículas, englobando hospitais universitários, museus e aposentadorias — e custo por aluno, que busca isolar apenas as despesas diretamente associadas ao ensino.

O mito de que os alunos das IES públicas custariam “duas a três vezes mais” que os das privadas foi difundido em relatórios do Banco Mundial (2017) e em

discursos políticos recentes, mas, segundo Bielschowsky e Amaral (2022), trata-se de uma interpretação equivocada. Aplicando metodologias adequadas, os autores demonstram que o custo anual médio por aluno em universidades federais, estaduais paulistas e confessionais privadas é bastante próximo, variando entre R\$ 18 mil e R\$ 19,7 mil em 2019. Em contraste, faculdades e centros universitários privados apresentam menores custos médios, em grande medida pela forte presença da educação a distância (EAD), modalidade que reduz despesas, mas gera preocupações sobre qualidade e permanência estudantil (Bielschowsky; Amaral, 2022).

Esse cenário evidencia um paradoxo: enquanto as IESs públicas, responsáveis por grande parte da pesquisa científica e pela oferta de cursos de maior complexidade, enfrentaram restrições orçamentárias severas — intensificadas pela Emenda Constitucional n. 95, de 15 de dezembro de 2016 —, bilhões de reais foram direcionados ao financiamento de matrículas privadas via Fies. Assim, a comparação entre setores não deve se limitar ao número de matrículas financiadas, mas sim considerar a eficiência relativa dos recursos aplicados e os resultados sociais obtidos.

Nesse mesmo sentido, Milan *et al.* (2016) ressaltam que diferentes estratégias de alocação de recursos — assim como estratégias de precificação nas empresas — resultam em diferentes níveis de desempenho, sendo essencial avaliar o equilíbrio entre custo, concorrência e valor entregue.

No plano regional, Santana *et al.* (2019) mostram que o financiamento público tende a se concentrar em regiões mais desenvolvidas, especialmente no Sudeste, ampliando desigualdades entre Norte e Nordeste. Essa constatação sugere que, tanto gastos diretos quanto indiretos, podem reforçar assimetrias estruturais caso não sejam acompanhados de políticas redistributivas.

Diante desse panorama, diversos autores defendem a necessidade de mensurar a eficiência do gasto público, traduzindo ineficiências em termos fiscais e sociais (Macedo; Rosadas, 2006). A DEA, ao estimar fronteiras de melhores práticas, permite verificar quanto de recursos poderia ser economizado ou quanto de produção adicional poderia ser gerado caso as instituições operassem em plena eficiência (Cooper; Seiford; Tone, 2007). Essa abordagem conecta diretamente a discussão sobre financiamento à eficiência operacional das IES brasileiras, objetivo central deste estudo.

2.4 ESTUDOS ANTERIORES

Diante da relevância do financiamento estudantil e de seus impactos sobre a sustentabilidade e o desempenho das IES, torna-se essencial compreender em que medida programas como Fies influenciam a eficiência operacional das instituições. Nesse contexto, a literatura empírica — nacional e internacional — tem recorrido amplamente à Análise Envoltória de Dados (DEA) e a técnicas complementares para mensurar a eficiência e a produtividade no ensino superior, oferecendo subsídios para a comparação entre diferentes modelos de financiamento e gestão.

A produção científica sobre eficiência em IES é vasta e tem se desenvolvido em múltiplos contextos, tanto no Brasil quanto no exterior. Desde o trabalho seminal de Charnes, Cooper e Rhodes (1978), a DEA consolidou-se como uma das principais ferramentas metodológicas para mensuração da eficiência educacional, dada sua capacidade de lidar simultaneamente com múltiplos insumos e produtos, além de permitir comparações relativas entre unidades de decisão.

Diversos estudos internacionais reforçam a relevância da DEA e de técnicas associadas, como o índice de produtividade de Malmquist e o bootstrap. Johnes (2006), por exemplo, avaliou 100 IES do Reino Unido, combinando DEA com bootstrap para medir eficiência técnica em cursos de graduação e pós-graduação. Os resultados indicaram alto nível de eficiência, e o uso do bootstrap permitiu definir intervalos de confiança robustos. De forma semelhante, Lee (2011) analisou 37 universidades australianas, integrando DEA-bootstrap com regressão de variáveis ambientais, e concluiu que a localização e a proporção de professores seniores influenciam positivamente a eficiência, enquanto o número de alunos impacta negativamente.

Na Europa, Parteka e Wolszczak-Derlacz (2013) compararam 266 IES de sete países (2001–2005), utilizando DEA-bootstrap e Malmquist. Observaram aumento médio anual de 4,5% na produtividade, sobretudo por mudanças de eficiência. Já Veiderpass e Mckelvey (2016), com base em 944 IES de 17 países, identificaram diferenças significativas entre a DEA clássica e a DEA-bootstrap, com destaque para a associação positiva entre eficiência e intensidade de pesquisa. Guccio, Martorana e Monaco (2016) avaliaram o impacto do Processo de Bolonha na Itália, constatando ganhos de eficiência e convergência entre instituições após as reformas. Barra e Zotti

(2016), ao analisar departamentos da Universidade de Salerno, reforçaram a influência das proxies de qualidade sobre os resultados de eficiência.

Estudos mais recentes incluem Andersson *et al.* (2017), que aplicaram DEA-bootstrap em 35 IES da Suécia, registrando ineficiência média de 12% e crescimento de produtividade de 1,7% ao ano. Navarro-Chavez, Gómez Monge e Torres Hernández (2017) examinaram 32 universidades públicas mexicanas e encontraram eficiência média de 0,77 (DEA-VRS), destacando limitações estruturais. Wolszczak-Derlacz (2017), em comparação entre Europa e EUA, observou alta ineficiência técnica, mas confirmou a importância de doações e economias de escala como determinantes de eficiência. Lehmann *et al.* (2018) verificaram que universidades italianas reagiram melhor à crise de 2008 do que as alemãs, sugerindo influência das estruturas de governança. Mais recentemente, Brzezicki (2020) comparou universidades públicas e privadas da Polônia, revelando maior eficiência das públicas em número de graduados, mas desempenho inferior dos egressos em termos salariais.

Estudos adicionais reforçam a importância de políticas públicas no desempenho das IES. Civera *et al.* (2020) analisaram a Iniciativa de Excelência em universidades alemãs e italianas, destacando efeitos positivos e negativos no desempenho da pesquisa. Martínez-CaMPILLO e Fernandez-Santos (2020) mostraram que universidades espanholas se tornaram mais eficientes durante a crise econômica de 2008. Agasisti *et al.* (2021) relacionaram a eficiência de 449 universidades russas ao crescimento econômico regional, demonstrando correlação estatística significativa. Guironnet e Peypoch (2018) exploraram a eficiência geográfica de universidades norte-americanas, ressaltando a importância da localização como fator de desempenho.

Nesse cenário, Nazarko e Šaparauskas (2014) avaliaram universidades públicas europeias, enfatizando a inclusão de variáveis qualitativas na análise. Mais recentemente, Arbula Blecich (2024) aplicou a DEA Window para 30 países europeus (2013–2020), identificando variações temporais significativas e defendendo a necessidade de análises dinâmicas.

No Brasil, a DEA foi amplamente utilizada para avaliar a eficiência das universidades públicas federais. Marinho, Resende e Façanha (1997) e Façanha e Marinho (1999) já discutiam a relação entre eficiência e modelos de financiamento das IES. Belloni (2000) propôs indicadores de eficiência baseados em qualidade da

graduação e pós-graduação, enquanto Façanha e Marinho (2001) ampliaram a análise para incluir IES privadas. Costa *et al.* (2012) aplicaram DEA-SBM e Malmquist a 49 IFES, identificando diferenças relevantes entre grupos de instituições com maior e menor envolvimento em pesquisa e pós-graduação.

Siqueira (2015) examinou o impacto do Programa Reuni (2007–2012) e constatou ganhos modestos de eficiência, atribuídos à ampliação de escala. Santos, Silva Júnior e Nunes (2022) analisaram 56 universidades federais (2012–2018) e encontraram eficiência média de 94,5% e aumento de produtividade de 2,2%, explicado pelo efeito catch-up. Silva e Rosa (2022) avaliaram 37 universidades federais (2014–2016), concluindo que mais da metade operava de forma eficiente, mas apontando a necessidade de ampliar os critérios de avaliação. Almeida *et al.* (2018), em estudo sobre departamentos de economia, identificaram queda significativa de eficiência técnica entre 2009 e 2012, associada a fatores socioeconômicos dos alunos.

Estudos recentes ampliaram a análise para o financiamento estudantil. Souza (2021) e Pontuschka (2016) demonstraram que o Fies ampliou o acesso ao ensino superior privado, mas gerou custos fiscais e riscos de sustentabilidade. Becker e Mendonça (2021) reforçaram esses achados ao apontar que, embora o programa tenha reduzido a evasão, aumentou o tempo médio de conclusão e elevou a inadimplência.

A análise crítica da literatura revela que a DEA se consolidou como a principal metodologia para mensuração da eficiência em IES, tanto em nível institucional quanto em políticas públicas de expansão. Os estudos internacionais destacam a importância de análises dinâmicas (DEA Window, Malmquist) e da incorporação de variáveis qualitativas, enquanto os estudos nacionais evidenciam avanços relevantes, mas ainda concentrados no setor público. Persiste, portanto, uma lacuna na comparação direta entre IES públicas e privadas e, sobretudo, na avaliação da eficiência relativa do Fies frente ao investimento direto nas IES públicas.

Para sistematizar os principais achados da literatura nacional e internacional sobre eficiência em Instituições de Ensino Superior, apresenta-se a Tabela 1, que resume o contexto das pesquisas, metodologias utilizadas e principais resultados encontrados. A síntese permite identificar padrões relevantes, como a predominância do uso da DEA, a incorporação crescente de técnicas dinâmicas (como o índice de Malmquist e o bootstrap) e a diferença de enfoques entre estudos internacionais —

com destaque para análises dinâmicas e variáveis qualitativas — e nacionais, ainda centrados principalmente em IES públicas e no papel de programas como o Reuni e o Fies.

Para sistematizar os principais achados da literatura nacional e internacional sobre eficiência em Instituições de Ensino Superior, apresenta-se o Quadro 1, que resume os principais grupos de estudos, metodologias utilizadas e contribuições identificadas na literatura. O quadro apresenta uma síntese analítica dos trabalhos mais representativos, enquanto a relação detalhada dos estudos revisados — incluindo contexto, amostra e resultados — encontra-se apresentada no Apêndice F. Essa organização permite evidenciar padrões relevantes da literatura, como a predominância do uso da Análise Envoltória de Dados (DEA), a incorporação crescente de técnicas dinâmicas (como o índice de Malmquist e o bootstrap) e as diferenças de enfoque entre estudos internacionais e nacionais.

Quadro 1 - Síntese da literatura empírica sobre eficiência em IES

Grupo de Estudos	Principais Autores	Abordagem Metodológica	Principais Contribuições
Estudos metodológicos seminais	Charnes, Cooper e Rhodes (1978)	DEA (CCR)	Introdução da Análise Envoltória de Dados para mensuração de eficiência com múltiplos inputs e outputs.
Estudos internacionais de eficiência em IES	Johnes (2006); Lee (2011); Parteka e Wolszczak-Derlacz (2013); Veiderpass e Mckelvey (2016); Wolszczak-Derlacz (2017); Brzezicki (2020); Agasisti et al. (2021); Arbula Blechich (2024)	DEA, DEA-bootstrap, DEA-Malmquist, DEA Window	Evidenciam ampla utilização da DEA para avaliação de eficiência institucional, com crescente incorporação de métodos dinâmicos e variáveis ambientais.
Estudos sobre determinantes institucionais e governança	Barra e Zotti (2016); Lehmann et al. (2018); Martínez-Campillo e Fernandez-Santos (2020); Guironnet e Peypoch (2018)	DEA com regressões de segunda etapa	Demonstram influência de fatores institucionais, regionais e de governança sobre os níveis de eficiência universitária.
Estudos sobre eficiência de universidades brasileiras	Marinho, Resende e Façanha (1997); Façanha e Marinho (1999, 2001); Belloni (2000); Costa et al. (2012); Siqueira (2015); Santos, Silva Júnior e Nunes (2022); Silva e Rosa (2022); Almeida et al. (2018)	DEA, DEA-SBM, Malmquist	Evidenciam evolução da eficiência das universidades federais e efeitos de políticas públicas como o Reuni.
Estudos sobre financiamento estudantil (Fies)	Pontuschka (2016); Souza (2021); Becker e Mendonça (2021)	Avaliações empíricas e qualitativas	Indicam expansão do acesso ao ensino superior privado, mas com custos fiscais e desafios de sustentabilidade.

Fonte: Elaboração própria com base na base de dados Web of Science.

Diante do panorama apresentado, observa-se que a literatura oferece contribuições relevantes para a compreensão dos conceitos de eficiência, produtividade e desempenho no ensino superior, bem como para a aplicação de métodos não paramétricos de avaliação, como a DEA. Contudo, os estudos analisados não respondem de forma conclusiva se os recursos públicos alocados por meio do financiamento estudantil indireto, via Fies, apresentam níveis de eficiência técnica comparáveis àqueles observados no modelo de oferta pública direta, materializado no financiamento das universidades federais.

Essa lacuna torna-se particularmente relevante sob a perspectiva do controle do gasto público e da avaliação de políticas públicas, ao envolver distintos arranjos institucionais de financiamento, com impactos orçamentários, sociais e educacionais diferenciados. Diante desse contexto, o presente estudo busca avançar na literatura ao investigar empiricamente os determinantes da eficiência técnica aplicada ao processo operacional das Instituições de Ensino Superior brasileiras, considerando tanto a dimensão estática quanto a dinâmica do desempenho institucional.

Com base nesse arcabouço teórico e empírico, formulam-se, a seguir, as hipóteses de pesquisa que orientam a análise empírica desenvolvida neste trabalho.

2.5 HIPÓTESES DE PESQUISA

A literatura revisada na seção anterior evidencia que a eficiência no ensino superior é influenciada por múltiplos fatores institucionais, organizacionais e contextuais, incluindo modelos de financiamento, características estruturais das instituições e condições socioeconômicas do ambiente em que estão inseridas. Conforme sintetizado no Quadro 1, diversos estudos nacionais e internacionais têm utilizado a Análise Envoltória de Dados (DEA) e métodos associados para investigar diferenças de eficiência entre instituições de ensino superior, bem como os determinantes institucionais e regionais desse desempenho.

No contexto internacional, pesquisas baseadas em DEA demonstram que diferenças estruturais entre instituições — incluindo modelos de financiamento, governança e estrutura acadêmica — podem produzir padrões distintos de eficiência e produtividade no ensino superior (Johnes, 2006; Parteka; Wolszczak-Derlacz, 2013; Veiderpass; Mckelvey, 2016). Resultados semelhantes são observados em análises comparativas entre sistemas universitários e entre tipos institucionais, evidenciando que arranjos organizacionais e políticas públicas influenciam diretamente o

desempenho das instituições (Wolszczak-Derlacz, 2017; Brzezicki, 2020; Agasisti et al., 2021).

No caso brasileiro, estudos empíricos indicam que as instituições públicas e privadas operam sob estruturas de financiamento e incentivos distintos, o que pode se refletir em diferenças nos níveis de eficiência observados (Façanha; Marinho, 2001; Costa et al., 2012; Santos; Silva Júnior; Nunes, 2022). Paralelamente, pesquisas sobre políticas de financiamento estudantil apontam que o Fies ampliou o acesso ao ensino superior privado, mas também introduziu novos incentivos e desafios relacionados à sustentabilidade financeira e ao desempenho institucional (Pontuschka, 2016; Souza, 2021; Becker; Mendonça, 2021).

Com base nesse arcabouço teórico e empírico, formulam-se as seguintes hipóteses de pesquisa.

2.5.1 Diferenças de eficiência entre modelos institucionais

A literatura indica que diferenças nos modelos de financiamento e nas estruturas organizacionais das instituições podem resultar em padrões distintos de eficiência no ensino superior. Estudos comparativos mostram que instituições públicas e privadas frequentemente apresentam níveis diferenciados de desempenho, refletindo variações em incentivos institucionais, estrutura de custos e objetivos organizacionais (Façanha; Marinho, 2001; Brzezicki, 2020; Santos; Silva Júnior; Nunes, 2022).

Diante dessas evidências, estabelece-se a seguinte hipótese:

H1: As Instituições de Ensino Superior privadas participantes do Fies apresentam padrões de eficiência técnica aplicada ao processo operacional sistematicamente distintos daqueles observados nas Instituições de Ensino Superior públicas federais.

2.5.2 Efeitos do financiamento estudantil sobre a eficiência institucional

A literatura sobre financiamento estudantil sugere que políticas públicas como o Fies podem alterar o comportamento institucional das IES privadas, ao modificar as condições de acesso, permanência e financiamento dos estudantes. Estudos empíricos indicam que o programa contribuiu para ampliar o acesso ao ensino

superior, mas também gerou mudanças na dinâmica de matrícula, evasão e conclusão dos cursos (Pontuschka, 2016; Souza, 2021; Becker; Mendonça, 2021).

Considerando que essas mudanças podem influenciar o desempenho operacional das instituições, formula-se a seguinte hipótese:

H2: A participação no Fies está associada a variações nos níveis de eficiência técnica aplicada ao processo operacional das Instituições de Ensino Superior privadas, mesmo após o controle por características institucionais, regionais e socioeconômicas.

2.5.3 Influência de fatores institucionais e regionais na eficiência

Diversos estudos internacionais demonstram que variáveis institucionais e contextuais desempenham papel relevante na explicação das diferenças de eficiência entre universidades. Fatores como localização geográfica, estrutura organizacional, intensidade de pesquisa e características regionais do mercado educacional têm sido identificados como determinantes importantes do desempenho institucional (Lee, 2011; Guironnet; Peypoch, 2018; Martínez-Campillo; Fernandez-Santos, 2020; Agasisti et al., 2021).

Com base nessas evidências, propõe-se a seguinte hipótese:

H3: Variáveis institucionais, regionais e socioeconômicas — tais como localização geográfica, porte institucional, tempo de credenciamento e contexto econômico local — influenciam significativamente os níveis de eficiência técnica aplicada ao processo operacional das Instituições de Ensino Superior.

2.5.4 Dinâmica temporal da produtividade institucional

A literatura internacional também evidencia a importância de análises dinâmicas da eficiência no ensino superior, frequentemente conduzidas por meio do índice de produtividade de Malmquist. Esses estudos indicam que as variações na produtividade das instituições tendem a resultar principalmente de mudanças na eficiência técnica e na escala de operação, enquanto deslocamentos tecnológicos na fronteira de produção ocorrem com menor frequência (Parteka; Wolszczak-Derlacz, 2013; Andersson et al., 2017; Arbula Blecich, 2024).

Com base nesse referencial empírico, estabelece-se a seguinte hipótese:

H4: As variações temporais da produtividade das Instituições de Ensino Superior decorrem predominantemente de ajustes na eficiência técnica e de escala, e não de deslocamentos persistentes da fronteira tecnológica.

As hipóteses formuladas orientam diretamente a estratégia empírica adotada nesta pesquisa. A hipótese H1 será investigada por meio da comparação dos escores de eficiência obtidos pela Análise Envoltória de Dados (DEA) entre Instituições de Ensino Superior públicas federais e instituições privadas participantes do Fies. A hipótese H2 será examinada por meio da análise da relação entre participação no Fies e os níveis de eficiência técnica estimados, considerando controles institucionais, regionais e socioeconômicos. A hipótese H3 será analisada a partir da incorporação de variáveis ambientais ao modelo empírico, permitindo identificar fatores contextuais associados às diferenças de eficiência entre instituições. Por fim, a hipótese H4 será avaliada por meio da decomposição do índice de produtividade de Malmquist, possibilitando distinguir as variações decorrentes de mudanças na eficiência técnica e de escala daquelas associadas a deslocamentos da fronteira tecnológica.

As hipóteses formuladas permitem avaliar, de forma integrada, o desempenho relativo das Instituições de Ensino Superior sob diferentes arranjos institucionais e mecanismos de financiamento, bem como investigar os determinantes estruturais e dinâmicos da eficiência técnica no ensino superior brasileiro, à luz das funções de planejamento, orçamento e controle do setor público.

A partir desse enquadramento analítico, procede-se à apresentação do desenho metodológico da pesquisa, das variáveis selecionadas e dos procedimentos empíricos adotados para mensurar e comparar a eficiência das diferentes modalidades de alocação de recursos públicos analisadas neste estudo.

3 MÉTODO DE PESQUISA

Este capítulo descreve a metodologia adotada para atender ao objetivo geral da pesquisa, que consiste em avaliar a eficiência técnica aplicada ao processo produtivo das IES brasileiras, públicas e privadas, no período de 2019 a 2023. O delineamento metodológico foi estruturado de modo a assegurar consistência científica, comparabilidade institucional e suporte à análise da eficiência e da produtividade no ensino superior brasileiro.

Do ponto de vista da controladoria pública, a abordagem metodológica adotada permite operacionalizar avaliações de desempenho institucional a partir de indicadores quantitativos de eficiência técnica, em linha com funções de monitoramento, avaliação e suporte à tomada de decisão no setor público. A utilização integrada de métricas de eficiência, produtividade e fatores condicionantes amplia a capacidade analítica dos sistemas de controle, ao permitir a identificação de assimetrias de desempenho e de possíveis desvios entre os recursos empregados e os resultados alcançados, aspecto central nas abordagens contemporâneas de avaliação de desempenho organizacional e de políticas públicas (Behn, 2003; Moynihan, 2008).

A opção por concentrar a análise nas Instituições de Ensino Superior públicas federais decorre de fundamentos jurídico-institucionais e de coerência metodológica. Embora alguns Estados mantenham e financiem suas próprias universidades estaduais, somente a União é simultaneamente responsável pelo financiamento indireto via Fies e pelo custeio direto de suas IES federais. Nesse sentido, a comparação entre gasto direto e financiamento estudantil indireto só é tecnicamente válida dentro do âmbito federal, uma vez que os Estados não participam do Fies como entes financiadores e possuem regimes orçamentários, contábeis e modelos de financiamento distintos daqueles da União. Assim, restringir o estudo às IES públicas federais assegura homogeneidade institucional, comparabilidade entre modalidades de financiamento e aderência ao objetivo central da pesquisa, evitando vieses decorrentes das profundas diferenças regulatórias, fiscais e operacionais existentes entre sistemas estaduais e federal de ensino superior.

Em relação ao segundo objetivo específico, procedeu-se à mensuração da eficiência operacional do Fies e das IES públicas por meio da DEA, metodologia não paramétrica desenvolvida por Charnes, Cooper e Rhodes (1978), a partir dos

fundamentos de Farrell (1957) e Shephard (1970). Foram utilizados modelos orientados a insumos com retornos variáveis de escala (DEA-BCC), após um teste de modelo, estimando-se índices de eficiência técnica e identificando-se outliers, retornos de escala e potenciais vieses de mensuração, de acordo com as recomendações metodológicas de Simar e Wilson (2007).

O terceiro objetivo específico buscou examinar a influência de variáveis contextuais sobre a eficiência das IES públicas e privadas por meio da aplicação de técnicas de aprendizado de máquina (ML). Diferentemente da abordagem tradicional baseada em regressões Tobit com correção por bootstrap (Simar; Wilson, 2007), foram empregados modelos capazes de lidar com relações não lineares e interações complexas entre variáveis. Para essa etapa, utilizaram-se três métodos principais de ML: Elastic Net, Random Forest e XGBoost. Entre as variáveis ambientais consideradas, destacam-se a localização geográfica, o tipo de rede, a modalidade de ensino, o tempo de existência e o PIB regional.

No cumprimento do quarto objetivo específico, realizou-se a comparação da eficiência dos gastos públicos destinados ao Fies com os investimentos diretos nas IES públicas federais. Essa etapa baseou-se na integração dos índices de eficiência e de produtividade obtidos nas análises anteriores, de modo a verificar em que medida o financiamento indireto via Fies se mostra eficiente na promoção do acesso, da permanência e da conclusão no ensino superior brasileiro (Souza, 2021; Becker; Mendonça, 2021; Pontuschka, 2016).

Por fim, o quinto objetivo específico foi contemplado com a aplicação do Índice de Produtividade de Malmquist, proposto por Caves, Christensen e Diewert (1982), que possibilita avaliar a evolução temporal da produtividade no ensino superior brasileiro entre 2019 e 2023. Esse procedimento metodológico permite distinguir ganhos decorrentes de melhorias gerenciais (catch-up effect) e avanços tecnológicos (frontier-shift effect), oferecendo uma perspectiva dinâmica sobre a eficiência das IES públicas e da política de financiamento via Fies.

Assim, a metodologia adotada combina rigor analítico e diversidade de técnicas, permitindo não apenas mensurar a eficiência relativa do Fies e das IES públicas, mas também compreender os fatores que a condicionam e refletir sobre a pertinência da alocação de recursos públicos no ensino superior. Em complemento, todos os cálculos decorrentes da aplicação das técnicas selecionadas são implementados por meio da linguagem de programação R, em cada uma das etapas

anteriormente descritas, detalhadas ao longo das seções subsequentes deste capítulo, garantindo padronização, reprodutibilidade e robustez às análises realizadas.

Os resultados de eficiência técnica e produtividade institucional obtidos neste estudo devem ser interpretados como instrumentos de apoio ao ciclo de políticas públicas no ensino superior, especialmente nas etapas de planejamento, alocação orçamentária, monitoramento e avaliação de resultados.

3.1 SELEÇÃO DE VARIÁVEIS

A seleção de variáveis constitui etapa central na modelagem do estudo de eficiência, uma vez que a adequação entre insumos e produtos influencia diretamente a robustez dos resultados obtidos pela DEA. Como destacam Coelli *et al.* (2005), a DEA é sensível ao número e à natureza das variáveis incluídas no modelo, de modo que a escolha deve refletir o processo produtivo analisado e preservar a comparabilidade entre as unidades decisórias (Decision Making Units – DMUs). No caso do ensino superior, essas unidades correspondem às IES, que transformam recursos humanos, financeiros e estruturais em resultados educacionais, científicos e sociais.

Na literatura especializada, observa-se ampla diversidade na definição de variáveis, embora haja convergência em torno de alguns elementos centrais. Entre os inputs mais utilizados destacam-se: número de docentes, número de técnicos administrativos, gastos correntes e investimentos, bem como número de matrículas, quando considerado insumo do processo educacional. Já os outputs mais recorrentes incluem número de concluintes, produção científica (artigos, livros, citações) e indicadores de qualidade, como o Índice Geral de Cursos (IGC) ou o Conceito Preliminar de Curso (CPC). As principais variáveis utilizadas em pesquisas anteriores estão sistematizadas no Quadro 2.

Além dos insumos e produtos que compõem o processo produtivo das IES, o presente estudo também incorpora variáveis ambientais. Essas variáveis representam fatores externos à gestão das instituições — como condições socioeconômicas, estruturas regionais e características institucionais — que influenciam os escores de eficiência, mas não são diretamente controlados pela administração. Entre elas incluem-se: PIB municipal, densidade demográfica, tipo de rede (pública/privada), tempo de credenciamento, localização da sede (capital ou interior), coordenadas

geográficas (latitude e longitude) e participação da IES no Fies. Esses elementos são frequentemente reconhecidos pela literatura como condicionantes importantes do desempenho institucional, justificando sua inclusão nas análises subsequentes.

Os dados utilizados neste estudo foram coletados em 2024, a partir de bases públicas disponibilizadas por órgãos oficiais de estatística e educação. As informações referentes aos insumos e produtos das Instituições de Ensino Superior foram extraídas das Sinopses Estatísticas da Educação (Brasil, [s.d.]c) e dos Microdados do Censo da Educação Superior (Brasil, [s.d.]c). Os indicadores de qualidade institucional, incluindo o Índice Geral de Cursos (IGC) (Brasil, [s.d.]b), foram obtidos no portal de dados abertos do Inep. As variáveis ambientais socioeconômicas — como Produto Interno Bruto (PIB) municipal (Brasil, [s.d.]e) e densidade demográfica (Brasil, [s.d.]d) — foram obtidas a partir das estatísticas oficiais do IBGE. Já a identificação da participação das IES no Fies foi realizada a partir das bases oficiais do Portal da Transparência da União, incluindo a execução orçamentária das ações relacionadas ao programa (Brasil, 2023b), e os registros específicos da Concessão de Financiamento Estudantil – Fies (Brasil, [s.d.]a) (Lei n. 10.260, de 2001), bem como das informações consolidadas pelo Fundo Nacional de Desenvolvimento da Educação – FNDE (Brasil, 2023a) sobre mantenedoras e IES habilitadas ao programa. Todas essas fontes garantem a confiabilidade, comparabilidade e transparência das informações utilizadas na modelagem.

Quadro 2 - Variáveis utilizadas em Estudos Anteriores

Variável	Tipo	Estudos Anteriores
Instituições de Ensino Superior - IES	DMU	Cao, Qiu e Huang (2022); Krejnuš <i>et al.</i> (2023); Santos <i>et al.</i> (2022); Siqueira (2015).
Docentes em exercício	Input	Arbula Blecich (2024); Cao, Qiu e Huang (2022); Johnes e Ruggiero (2016); Moncayo-Martínez <i>et al.</i> (2020); SANTOS <i>et al.</i> (2022); Siqueira (2015).
Funcionários técnico-administrativos	Input	Arbula Blecich (2024); Cao, Qiu e Huang (2022); Guironnet e Peypoch (2018); Nishimura e Okamuro (2016); Santos <i>et al.</i> (2022).
Matrículas totais	Input	Arbula Blecich (2024); Moncayo-Martínez <i>et al.</i> (2020).
Infraestrutura digital de suporte acadêmico	Input	Arbula Blecich (2024); Abramo <i>et al.</i> (2015).
Acervo das bibliotecas	Input	Nazarko e Šaparauskas (2014); Moncayo-Martínez <i>et al.</i> (2020).
Concluintes	Output	Contreras e Lozano (2020); Gul e Yucesan (2022); Guironnet e Peypoch (2018); Nishimura e Okamuro (2016); Santos <i>et al.</i> (2022); Silva e Rosa (2022).
Evasão/Desvinculação	Output	Bean (1980).
Estudantes beneficiados por programas de inclusão	Input	Agasisti e Pérez-Esparrells (2010).

Índice Geral de Cursos (IGC)	Output	Bittencourt <i>et al.</i> (2009).
Produto Interno Bruto Total	Ambiental	Arbula Bleich (2024).
Densidade Demográfica	Ambiental	Afonso <i>et al.</i> (2005)
Rede de Ensino	Ambiental	Santos <i>et al.</i> (2022)
Ano de Credenciamento da IES	Ambiental	Siqueira (2015).
Sede da IES na Capital	Ambiental	Pereira (2023).
Latitude e Longitude da IES	Ambiental	Pereira (2023).
Participação da IES no FIES	Ambiental	Souza (2021); Pontuschka (2016); Becker e Mendonça (2021).

Fonte: Dicionário de Dados (Brasil, [s.d.]).

O Quadro 2 apresenta a descrição e a metodologia das variáveis selecionadas para o estudo — abrangendo inputs, outputs e variáveis ambientais — detalhando sua mensuração a partir das bases oficiais do Inep, IBGE e FNDE. Essa sistematização permite identificar a aderência das variáveis ao processo produtivo do ensino superior, bem como sua fundamentação teórica e disponibilidade nas fontes de dados.

Ao combinar variáveis internas e externas ao processo produtivo, busca-se construir um modelo mais abrangente e fiel à realidade das IES brasileiras. Essa integração assegura que os escores de eficiência obtidos pela DEA reflitam tanto o desempenho operacional das instituições quanto as condições estruturais e contextuais que influenciam seu funcionamento, fortalecendo a validade analítica e interpretativa dos resultados.

Quadro 3 - Variáveis escolhidas – Descrição e Metodologia

Variável	Tipo	Descrição e Metodologia
Instituições de Ensino Superior - IES	DMU	Unidade de análise utilizada para avaliar a eficiência relativa de entidades comparáveis, definida neste estudo como as Instituições de Ensino Superior (IES), a partir dos dados de identificação disponíveis.
Docentes em exercício	Input	Quantidade total de docentes em atividade nas IES, obtida a partir da variável QT_DOC_EXE.
Funcionários técnico-administrativos	Input	Número total de servidores técnico-administrativos vinculados às IES, mensurado pela variável QT_TEC_TOTAL.
Matrículas totais	Input / Output	Total de estudantes matriculados nas IES em cada ano do período analisado, captado pela variável QT_MAT, com distinção entre matrículas presenciais e a distância (EAD).
Infraestrutura digital de suporte acadêmico	Input	Conjunto de recursos e serviços digitais oferecidos pelas IES para apoio à pesquisa e ensino, consolidado pelas variáveis IN_ACESSO_PORTAL_CAPES, IN_ACESSO_OUTRAS_BASES, IN_ASSINA_OUTRA_BASE, IN_REPOSITARIO_INSTITUCIONAL, IN_BUSCA_INTEGRADA e IN_PARTICIPA_REDE_SOCIAL.

Acervo das bibliotecas	Input	Indicador do suporte acadêmico físico e digital disponível nas IES, consolidado pelas variáveis QT_PERIODICO_ELETRONICO e QT_LIVRO_ELETRONICO.
Concluintes	Input / Output	Número total de estudantes concluintes registrados no ano-base, aferido pela variável QT_CONC, discriminando concluintes das modalidades presencial e EAD.
Evasão/Desvinculação	Output (Indesejável)	Número de estudantes que, no ano-base, se desvincularam da IES por abandono, falecimento ou outros motivos que resultaram na não conclusão do curso.
Estudantes beneficiados por programas de inclusão	Output	Quantidade de estudantes matriculados contemplados por políticas de inclusão, como o ProUni, o Fies e as cotas sociais e raciais.
Índice Geral de Cursos (IGC)	Output	Indicador de desempenho acadêmico das IES, calculado a partir da média ponderada das notas dos cursos de graduação e pós-graduação, conforme avaliações oficiais realizadas no ano-base.
Produto Interno Bruto Total	Ambiental	PIB Total correspondente ao total dos PIB de cada município, segundo dados do IBGE.
Densidade Demográfica	Ambiental	Densidade Demográfica de cada município, considerando as populações e dimensões territoriais de cada município que as compõe.
Rede de Ensino	Ambiental	Tipo de Rede de Ensino da IES: 1 – Pública; 2 – Provada.
Ano de Credenciamento	Ambiental	Ano de credenciamento da IES junto ao MEC como forma de mensurar o tempo de existência da instituição.
Sede na Capital	Ambiental	Identifica se a sede da IES se encontra na capital do estado.
Latitude e Longitude	Ambiental	A localização geográfica é um fator relacionado com a eficiência e a produtividade.
Indicador de participação da IES no Fies	Ambiental	Variável dummy binária (0/1) que identifica se a instituição possui matrículas financiadas pelo Fies no ano t ; assume valor 1 quando há ao menos uma matrícula financiada e 0, caso contrário. Fonte: Censo da Educação Superior/Inep e FNDE.

Fonte: Dicionário de Dados (Brasil, [s.d.]).c).

Entre os outputs selecionados, merece destaque o Índice Geral de Cursos (IGC), instituído em 2007 e divulgado anualmente pelo Inep a partir de 2008. O IGC foi concebido como um indicador sintético de qualidade institucional, combinando a média ponderada dos Conceitos Preliminares de Curso (CPC) da graduação e os conceitos atribuídos pela Capes aos programas de pós-graduação *stricto sensu*. Sua metodologia considera a proporção de matrículas na graduação, mestrado e doutorado, de modo a refletir o peso relativo de cada nível de ensino na composição final da nota (BRASIL, [s.d.]).b). O resultado é um índice contínuo, variando de 0 a 5, que sintetiza o desempenho institucional em diferentes dimensões acadêmicas. Apesar de sua relevância como medida oficial, a literatura alerta para o uso do IGC também como instrumento de marketing e ranqueamento (Bittencourt, 2009), o que

pode ampliar seu impacto simbólico, mas nem sempre traduzir integralmente a qualidade educacional. Neste estudo, o IGC foi incorporado como variável de output, representando o desempenho acadêmico agregado das Instituições de Ensino Superior.

A análise dessa sistematização evidencia que há padrões consolidados na literatura: a utilização de recursos humanos e financeiros como insumos e a consideração de resultados acadêmicos, de pesquisa e de qualidade como produtos. Ademais, alguns estudos incorporam variáveis adicionais, como taxas de evasão, depósitos de patentes ou participação de alunos em programas sociais, refletindo especificidades contextuais.

No presente estudo, a seleção de variáveis foi orientada por três critérios principais: a) relevância teórica, ou seja, aderência ao processo produtivo das IES; b) disponibilidade de dados consistentes e comparáveis em fontes oficiais, como o Censo da Educação Superior (Brasil, [s.d.]) e indicadores do Sinaes; e c) aderência metodológica, assegurando que o número de variáveis não comprometa o poder discriminatório da DEA (Cooper; Seiford; Zhu, 2011). Dessa forma, foram escolhidos inputs representativos da estrutura institucional, como corpo docente, corpo técnico e número de matrículas, e outputs relacionados aos resultados acadêmicos, incluindo concluintes e indicadores de desempenho institucional.

Ao privilegiar variáveis amplamente reconhecidas na literatura e disponíveis em bases públicas de dados, buscou-se assegurar comparabilidade com estudos anteriores e robustez analítica, ao mesmo tempo em que se preserva a especificidade do contexto brasileiro e a pertinência da comparação entre IES públicas e o programa Fies. A etapa seguinte aprofunda a discussão sobre a identificação de outliers e sua relevância para a precisão dos escores de eficiência.

3.2 IDENTIFICAÇÃO DE OUTLIERS E SUPEREFICIÊNCIA

A identificação de outliers é uma etapa fundamental para a robustez da DEA, uma vez que valores extremos são observações em um conjunto de dados que se encontram a uma distância anormalmente grande de outros valores, podem ser registros incorretos ou não e distorcer a fronteira eficiente e comprometendo a validade dos escores de eficiência. Como destacam Cooper, Seiford e Tone (2007), o tratamento de observações discrepantes pode assumir diferentes estratégias: a) exclusão de outliers; b) substituição pelos limites do intervalo interquartil (IQR); ou

c) manutenção dos valores quando representarem características intrínsecas das DMUs. Neste estudo, adotou-se uma combinação de métodos estatísticos, análise visual por boxplots e técnicas complementares, de forma a preservar a consistência e a comparabilidade da amostra.

Entre os procedimentos empregados, utilizou-se a técnica de supereficiência, proposta por Andersen e Petersen (1993), capaz de identificar unidades extremamente eficientes, e a análise de superineficiência, que evidencia DMUs com desempenho significativamente abaixo da fronteira. Essas abordagens permitem sinalizar casos que merecem tratamento específico, evitando vieses na estimação dos resultados. De modo complementar, aplicaram-se técnicas de bootstrap, conforme Bogetoft e Otto (2011), para validar a robustez dos escores de eficiência obtidos.

Para assegurar a qualidade dos dados utilizados no pacote deaR, alguns ajustes prévios foram necessários. Valores iguais a zero, identificados em variáveis de input e output, foram substituídos por uma constante mínima (0,001). Segundo Bowlin (1998), essa prática evita distorções no cálculo sem alterar significativamente a fronteira de eficiência. Além disso, aplicou-se a transformação logarítmica natural às variáveis (equação 4), o que reduziu discrepâncias de magnitude e suavizou distribuições assimétricas, contribuindo para maior estabilidade computacional do modelo (Bogetoft; Otto, 2011).

$$x' = \ln(1 + x) \quad (4)$$

em que:

x : valor original da variável;

x' : valor transformado;

$\ln()$: logaritmo natural;

+1: constante somada para evitar a aplicação do logaritmo em valores iguais a zero.

Após a transformação, as variáveis foram normalizadas no intervalo [0,1] (equação 5), conforme proposto por Charnes, Cooper e Rhodes (1978), garantindo comparabilidade entre diferentes escalas e unidades de medida.

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)} \quad (5)$$

em que:

x' : valor normalizado;

$\min(x)$: menor valor de x na variável;

$\max(x)$: maior valor de x na variável.

O tratamento de outliers baseou-se no critério do intervalo interquartil (IQR), em que valores acima de $Q3+1,5 \times IQR$ ou abaixo de $Q1-1,5 \times IQR$ foram classificados como discrepantes (Cook; Zhu, 2007).

$$IQR=Q3-Q1 \quad IQR = Q3 - Q1 \quad IQR=Q3-Q1 \quad (6)$$

em que:

Q3: terceiro quartil (75%);

Q1: primeiro quartil (25%).

Os limites inferior e superior foram definidos como:

$$\text{Limite Inferior} = Q1 - 1,5 \times IQR \quad (7)$$

$$\text{Limite Superior} = Q3 + 1,5 \times IQR \quad (8)$$

Os valores que se situaram fora desse intervalo foram classificados como outliers, por apresentarem comportamento atípico em relação ao padrão observado na amostra.

Três formas de tratamento de outliers foram consideradas: a) remoção da unidade da amostra; b) substituição do valor pelo limite inferior ou superior do intervalo interquartil; e c) manutenção do valor original, nos casos em que representava uma característica inerente e real da DMU. Considerando a segurança e a confiabilidade das fontes de dados utilizadas, optou-se por manter os outliers na base, preservando a integridade das informações. No capítulo específico dedicado à análise de robustez, será apresentado um comparativo entre as estimativas produzidas com e sem a exclusão de outliers, permitindo avaliar o impacto dessas observações extremas sobre os escores de eficiência. Essa etapa também contará com funções implementadas em R e inspeção visual por boxplots para justificar metodologicamente a opção adotada.

Adicionalmente, aplicou-se a técnica de fronteira invertida, proposta por Yamada, Matui e Sugiyama (1994) e aprimorada por Entani, Maeda e Tanaka (2002), que consiste em inverter inputs e outputs para a construção de uma fronteira pessimista. Essa abordagem possibilita identificar IES superineficientes, ou seja, unidades com desempenho sistematicamente inferior à média do conjunto analisado.

Os resultados obtidos a partir dessas técnicas auxiliaram na detecção de potenciais erros de mensuração e na identificação de casos atípicos de desempenho elevado ou de baixa performance relativa. Cabe ressaltar, entretanto, que a decisão

sobre o tratamento de cada outlier não é puramente mecânica. Como observam Simar e Wilson (2015), há sempre certo grau de subjetividade envolvido, de modo que valores extremos podem refletir inovações tecnológicas ou práticas de gestão diferenciadas que merecem ser preservadas no conjunto de dados.

3.3 DEA-CCR E DEA-BCC

A Análise Envoltória de Dados (DEA), proposta inicialmente por Charnes, Cooper e Rhodes (1978), é um método não paramétrico amplamente utilizado para medir a eficiência relativa de unidades produtivas homogêneas, denominadas Decision Making Units (DMUs). O primeiro modelo, conhecido como DEA-CCR, baseia-se na hipótese de retornos constantes de escala (CRS), pressupondo que variações proporcionais nos insumos resultam em variações proporcionais nos produtos. Tal abordagem é adequada em cenários onde as unidades operam em escala ótima de produção, ou seja, sem ganhos ou perdas referentes ao porte (Charnes; Cooper; Rhodes, 1978; Coelli *et al.*, 2005).

Posteriormente, Banker, Charnes e Cooper (1984) ampliaram a técnica, originando o DEA-BCC, que relaxa a hipótese de proporcionalidade estrita e admite retornos variáveis de escala (VRS). Essa formulação permite identificar três regimes distintos: a) retornos crescentes de escala, em que aumentos proporcionais de insumos resultam em aumentos mais que proporcionais de outputs; b) retornos decrescentes de escala, quando os incrementos de insumos geram acréscimos menores nos produtos; e c) retornos constantes de escala, quando a proporcionalidade se mantém.

A principal contribuição do modelo BCC em relação ao CCR é possibilitar a decomposição da eficiência técnica global em dois componentes: eficiência técnica pura (ETP), que reflete a utilização dos recursos independentemente do porte da DMU, e eficiência de escala (EE), que avalia o quão próximo o tamanho da unidade está de sua escala ótima de operação (Farrell, 1957; Rosano-Peña; Gomes, 2018). Essa decomposição é particularmente relevante no contexto educacional, em que as instituições apresentam grande heterogeneidade em estrutura, porte e condições de operação (Nazarko; Šaparauskas, 2014; Siqueira, 2015).

Assim, a utilização conjunta dos modelos DEA-CCR e DEA-BCC fornece uma visão mais abrangente da eficiência, permitindo distinguir se a ineficiência observada

decorre do mau uso dos insumos ou do porte inadequado da DMU (Banker; Charnes; Cooper, 1984; Pereira, 2023).

No presente estudo, ambos os modelos foram empregados, pois as Instituições de Ensino Superior (IES) e o Fies apresentam naturezas distintas e não podem ser analisados sob a premissa de escala ótima uniforme. Dessa forma, a comparação entre CCR e BCC permite não apenas avaliar o desempenho relativo, mas também compreender de que forma o porte institucional e a escala de operação influenciam os resultados.

Essa combinação é ótima quando os valores de eficiência são maximizados em relação à fronteira de eficiência. Há duas formas básicas para a aplicação desse método. A primeira delas é a ideia desenvolvida com base nos trabalhos de Shephard e Farrell aperfeiçoado pelo método não paramétrico de Charnes, Cooper e Rhodes em 1978. Esse método, também conhecido como CCR – ou CRS, do inglês *Constant Return to Scale* – parte da hipótese da existência de um retorno constante de escala, quando é esperado um comportamento proporcional entre insumos e produtos. Matematicamente, o Problema de Programação Linear (PPL) orientados aos *inputs*, pode ser assim descrito:

$$\max h_o = \sum_{j=1}^m u_j y_{jo} \quad (9)$$

sujeito a:

$$\begin{aligned} \sum_{i=1}^n v_i x_{io} &= 1 \\ \sum_{j=1}^m u_j y_{jk} &\leq \sum_{i=1}^n v_i x_{ik}, \quad k = 1, 2, 3, \dots, n \\ u_i, v_i &\geq 0, \forall i, j \end{aligned}$$

Nesse modelo, v_i e u_j correspondem aos pesos atribuídos aos insumos (x_i) e produtos (y_j), definidos pela programação linear. Tais pesos podem ser interpretados como preços-sombra na fronteira eficiente, em que u_j indica o valor marginal do produto e v_i o custo marginal do insumo, refletindo a taxa de troca entre inputs e outputs. O escore de eficiência h_o da DMU_o é limitado superiormente por 1.

Um método semelhante foi desenvolvido com o PPL com retornos constantes de escala (CRS) orientado aos produtos:

$$\min \phi_0 = \sum_{i=1}^n v_i X_{io} \quad (10)$$

sujeito a:

$$\begin{aligned} \sum_{i=1}^n u_i Y_{io} &= 1 \\ \sum_{j=1}^m u_j y_{jk} &\geq \sum_{i=1}^n v_i x_{ik}, \quad k = 1, 2, 3, \dots, n \end{aligned}$$

$$u_j, v_i \geq 0, \forall i, j$$

Novamente, para esse caso, v_i e u_j indicam os pesos ponderados (para os insumos e produtos respectivamente) gerados pela programação linear de modo a existir uma combinação ótima entre os inputs e os outputs para a máxima eficiência. O valor obtido quando orientado aos produtos (ϕ_0) é maior ou igual a 1, de forma que $h_0 = 1/\phi_0$.

O outro método para a utilização do DEA BCC-io inclui fatores de escala (v_* e u_*). Esse cálculo é descrito pela seguinte formulação de um PPL variável orientado aos insumos:

$$\min \psi_0 = \sum_{i=1}^n v_i X_{i0} + v_* \quad (11)$$

sujeito a:

$$\begin{aligned} \sum_{j=1}^m u_j Y_{j0} &= 1; \\ \sum_{i=1}^n v_i X_{il} + k - \sum_{j=1}^m v_j X_{jl} &\geq 0, \\ u_j, v_i &\geq 0, \forall i, j; v_* - \text{irrestrito. } i = 1, \dots, n.; j = 1, \dots, m; l = 1, \dots, N. \end{aligned}$$

Quando utilizado o mesmo método, porém com a orientação ao produto, os termos da equação passam a ser os seguintes:

$$\max \omega_0 = \sum_{j=1}^m u_j Y_{j0} + u_* \quad (12)$$

sujeito a:

$$\begin{aligned} \sum_{i=1}^n v_i X_{i0} &= 1; \\ \sum_{i=1}^n u_j y_{jl} + k - \sum_{i=1}^n v_i X_{il} &\leq 0, \\ u_j, v_i &\geq 0, \forall i, j; u_* - \text{irrestrito. } i = 1, \dots, n.; j = 1, \dots, m; l = 1, \dots, N. \end{aligned}$$

Nos dois PPLs anteriores, um orientado aos insumos e o outro orientado aos produtos, v_* e u_* indicam os fatores de escala.

No modelo orientado aos **inputs**, quando v_* é positivo, indica retornos crescentes de escala; quando negativos, indica retornos decrescentes de escala; caso sejam nulos, indica retornos constantes de escala. Já no modelo orientado aos **outputs**, quando u_* é positivo, indica retornos decrescentes de escala; quando negativos, indicam retornos crescentes de escala; caso sejam nulos, é de retornos constantes de escala (Meza *et al.*, 2005).

Como nos modelos CRS, aqui, o valor obtido quando orientado aos produtos (ω_0) é maior ou igual a 1. Já no caso dos insumos (ψ_0), o valor é delimitado para ficar entre 0 e 1. Com isso, tanto o modelo CCR quanto o BCC podem ser analisados pela

ótica da orientação para insumos ou pela orientação para o produto de acordo com a análise a ser empregada.

Outro ponto a ser observado é como esses modelos se relacionam. Como já discutido, a eficiência técnica pode ser decomposta em eficiência técnica pura e de escala. A primeira é dada pelo cálculo DEA-BCC para cada DMU_i. A segunda é calculada a partir da razão entre a eficiência DEA-CCR dividida pela eficiência DEA-BCC.

Dessa forma, a ineficiência de escala surge da ideia de que a escala de produção pode afetar a eficiência. Em outros termos, há uma escala ótima de conversão de insumos em produtos de acordo com a máxima produtividade, de forma que quaisquer alterações para mais ou para menos tornam a unidade ineficiente (Mariano, 2005; Rosano-Peña, 2008; Boueri *et al.*, 2015).

Convém ressaltar que, em trabalhos relacionados à educação, a definição do modelo a ser adotado costuma ocorrer de maneira arbitrária pelos autores, seja em função do objetivo específico da pesquisa, seja pela opção de replicar metodologias já utilizadas em estudos anteriores. No presente estudo, entretanto, a escolha do modelo não é feita de forma aleatória: adota-se um procedimento de teste comparativo para verificar qual configuração é mais adequada a cada caso, se a de retornos constantes de escala (CRS) ou a de retornos variáveis de escala (VRS).

Adicionalmente, cabe destacar que este trabalho calcula os índices de eficiência tanto para o modelo de retornos constantes quanto para o de retornos variáveis, utilizando a orientação a produtos. Essa escolha se justifica porque, em avaliações de programas educacionais, a preocupação central reside em maximizar os resultados obtidos. Por fim, inclui-se também a análise da ineficiência de escala, de modo a fornecer uma visão mais abrangente sobre o desempenho das unidades avaliadas.

3.3.1 Network DEA

O modelo Network DEA (NDEA) surgiu como uma extensão da Análise Envoltória de Dados tradicional, com o objetivo de lidar com sistemas produtivos complexos que apresentam subprocessos interdependentes. Ao contrário dos modelos clássicos CCR e BCC, que tratam cada unidade de decisão (DMU) como uma “caixa-preta”, relacionando diretamente inputs e outputs, o NDEA permite decompor o processo produtivo em etapas ou redes internas, considerando fluxos

intermediários que conectam os subprocessos (Färe; Grosskopf, 2000; Tone; Tsutsui, 2009).

Essa característica torna o modelo particularmente adequado para avaliar a eficiência em ambientes educacionais, nos quais os resultados finais (formação de concluintes e qualidade institucional) dependem de etapas intermediárias, como alocação de recursos, processos de ensino e permanência estudantil (Lewis; Sexton, 2004; Pereira, 2023). Assim, o NDEA possibilita capturar melhor as relações de causalidade entre diferentes dimensões de desempenho das IES e, no caso deste estudo, comparar também os efeitos das políticas de financiamento estudantil via Fies.

Outra vantagem do modelo é que ele possibilita analisar tanto a eficiência global da DMU quanto a eficiência parcial de cada subprocesso, permitindo identificar em qual etapa se encontram as maiores fragilidades ou potenciais ganhos (Kao, 2014). Essa decomposição é especialmente útil para orientar políticas públicas, pois evidencia se a ineficiência está concentrada, por exemplo, no processo de acesso, de permanência ou nos resultados finais de formação.

No presente estudo, a aplicação do Network DEA foi organizada em três etapas que representam a lógica de funcionamento das IES e do Fies:

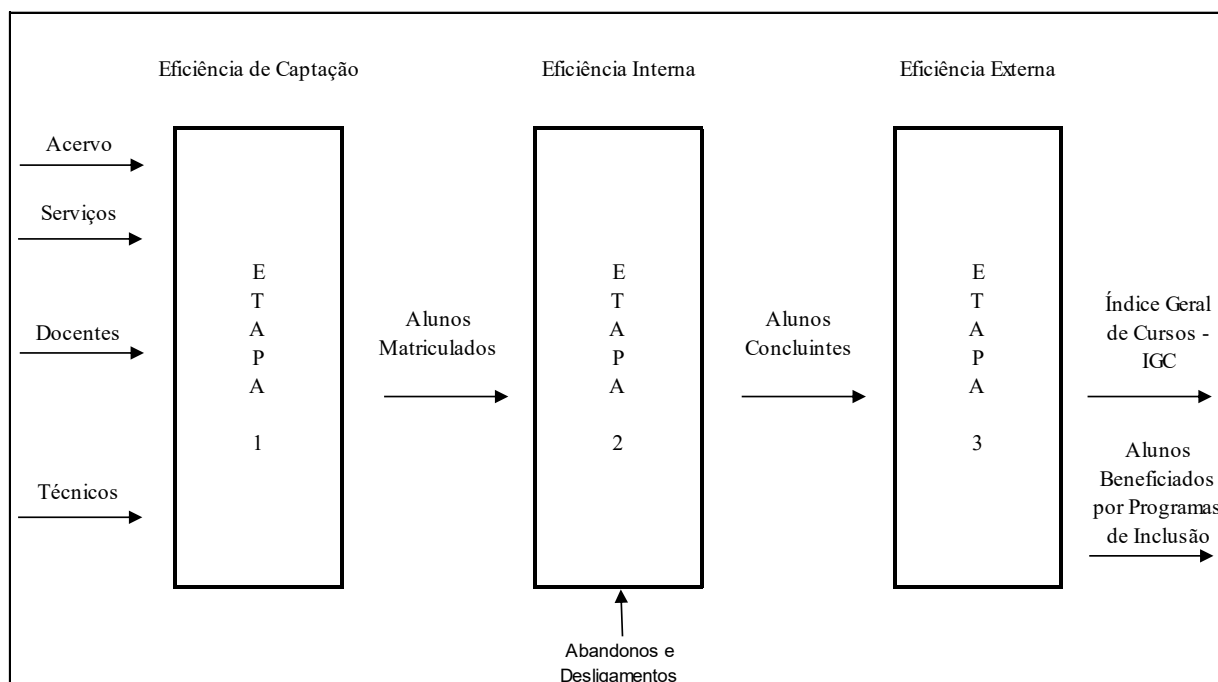
Etapa 1 – Eficiência de Captação: envolve a relação entre docentes e técnicos administrativos, além dos serviços ofertados e acervo disponibilizado, em comparação com o número de alunos efetivamente matriculados.

Etapa 2 – Eficiência Interna: corresponde à transformação das matrículas em concluintes, considerando também a ocorrência de abandono, que é tratado como uma externalidade indesejada no processo acadêmico.

Etapa 3 – Eficiência Externa: avalia os resultados finais alcançados pelas instituições a partir da avaliação dos cursos ofertados pelo IGC e da capacidade de promover inclusão social por meio de programas direcionados a esse fim.

No caso específico do presente trabalho, já considerando as variáveis utilizadas em cada uma das três etapas, a distribuição dos inputs e outputs é dada conforme a Figura 5.

Figura 5 - Realização do cálculo DEA em três etapas (Network DEA)



Fonte: Elaboração própria.

Essa abordagem permite não apenas verificar a eficiência global, mas também distinguir se os recursos são aplicados de forma equilibrada em cada etapa da cadeia produtiva do ensino superior.

3.4 TESTE DE MODELOS

Após o conhecimento de cada uma das possíveis orientações (insumos ou produtos) assim como os métodos possíveis de DEA (CCR ou BCC), a etapa seguinte é identificar o modelo mais apropriado a ser adotado neste estudo. Esse teste é importante porque a discricionariedade na escolha de um modelo (CCR ou BCC) ou a adoção dos dois modelos simultaneamente pode não ser apropriada. De acordo com Simar e Wilson (2002), impor, *a priori*, sem o teste apropriado, uma conjectura sobre os retornos de escala pode distorcer seriamente os índices de eficiência, principalmente, se o verdadeiro tipo de fronteira for outro. Desse modo, no desenvolvimento e na validação do método, deve-se testar o tipo de fronteira indicada pelos dados.

De acordo com Coelli *et al.* (2005), a diferença entre as medidas de eficiência calculadas pelos modelos CRS e VRS indica diretamente o nível de eficiência de escala das unidades. Assim, a escolha entre CCR e BCC deve ser fundamentada em critérios estatísticos e na adequação ao objeto de análise.

Segundo Simar e Wilson (2002), a seleção do modelo pode ser formalizada por meio do seguinte par de hipóteses:

- Hipótese nula (H_0): a fronteira tecnológica segue o modelo CCR (retornos constantes).
- Hipótese alternativa (H_a): a fronteira tecnológica segue o modelo BCC (retornos variáveis).

Caso H_0 seja verdadeira, os escores de eficiência obtidos sob BCC não diferirão estatisticamente dos calculados sob CCR. Contudo, se os escores CCR forem sistematicamente menores, a hipótese alternativa deve ser aceita, indicando a presença de retornos variáveis de escala.

A estatística de teste é construída a partir da razão entre as eficiências dos modelos BCC e CCR para cada DMU. A medida de eficiência de escala (S) orientada a insumos pode ser definida como:

$$S = \sum_{l=1}^n \frac{E_{BCC}^l}{E_{CCR}^l}, (k = 1, 2, 3, \dots, n) \quad (13)$$

Como $S \leq 1$, valores próximos de 1 indicam que a hipótese de retornos constantes não deve ser rejeitada, enquanto valores significativamente inferiores a 1 sugerem a rejeição de H_0 e a adoção do modelo BCC. Normalmente, aplica-se um nível de significância de 5% ($p \leq 0,05$) para definir o ponto de rejeição.

No presente estudo, para avaliar a adequação entre os modelos, utilizou-se o teste de Kolmogorov-Smirnov (KS), disponível no pacote stats do R. Esse teste compara duas distribuições empíricas — no caso, os conjuntos de eficiências dos modelos CCR e BCC — a fim de verificar se podem ser consideradas provenientes da mesma distribuição. A hipótese nula do teste assume que ambas são estatisticamente equivalentes, sendo a decisão de rejeição ou não tomada a partir do valor de p obtido. O teste pode ser aplicado de forma bilateral (“*two.sided*”) ou unilateral (“*less*”/“*greater*”), tendo como estatística de teste a máxima diferença entre as funções de distribuição acumulada empíricas das duas amostras (Feller, 1948; Conover, 1971; Marsaglia, Tsang e Wang, 2003).

A aplicação desse procedimento fornece os elementos necessários para avaliar a adequação entre os modelos CCR e BCC em termos estatísticos, respeitando a heterogeneidade do setor de ensino superior. A análise dos resultados será aprofundada no capítulo seguinte, que contempla não apenas a aplicação de técnicas de bootstrap para reforçar a robustez dos achados, mas também a utilização de testes

de comparação entre grupos — em especial, o teste de Kruskal-Wallis — de modo a investigar possíveis diferenças de eficiência entre IES públicas e privadas.

3.5 ANÁLISE DE GRUPOS

A análise de grupos é um recurso metodológico relevante quando se busca identificar se diferentes categorias de unidades de decisão (DMUs) apresentam desempenhos estatisticamente distintos em termos de eficiência. No contexto do ensino superior, a comparação entre Instituições de Ensino Superior (IES) públicas e privadas possibilita investigar se as diferenças estruturais entre essas redes — relacionadas ao porte, financiamento, perfil acadêmico e recursos disponíveis — impactam os resultados de eficiência obtidos no modelo de Análise Envoltória de Dados em Rede (NDEA).

Para essa finalidade, adota-se o teste de Kruskal-Wallis (KW), que se apresenta como uma alternativa não paramétrica à ANOVA. Esse teste é equivalente ao Mann-Whitney e apropriado para a comparação de grupos quando não se pode assumir a normalidade na distribuição dos dados — condição comum em escores de eficiência, que tendem a ser assimétricos e limitados (censurados) ao intervalo $[0,1]$ (Conover, 1998).

O procedimento baseia-se na ordenação (ranks) dos valores de eficiência e na comparação entre as somas desses ranks para os grupos analisados. A hipótese nula (H_0) do teste estabelece que as distribuições de eficiência entre os grupos são estatisticamente iguais, ao passo que a hipótese alternativa (H_a) sugere a existência de diferenças significativas em pelo menos um grupo. Quando o valor do p obtido é inferior ao nível de significância (comumente 5%), rejeita-se H_0 e conclui-se pela existência de diferenças entre as distribuições de eficiência.

No presente estudo, o teste de Kruskal-Wallis é aplicado aos escores do NDEA estático, abrangendo as diferentes etapas do modelo (Eficiência de Captação, Eficiência Interna e Eficiência Externa). Dessa forma, busca-se verificar se as IES públicas e privadas apresentam padrões distintos de eficiência em cada etapa do processo produtivo do ensino superior, considerando-se as particularidades de estrutura e gestão de cada rede.

Após a apresentação do teste de Kruskal-Wallis para comparação entre grupos, o próximo passo consiste na utilização das técnicas de bootstrap, que acrescentam robustez e confiabilidade às estimativas de eficiência.

3.6 BOOTSTRAP

A aplicação da DEA apresenta limitações relacionadas à sua natureza determinística, uma vez que todos os desvios em relação à fronteira eficiente são atribuídos à ineficiência das unidades de decisão (DMUs), sem considerar a possibilidade de erros amostrais ou variações aleatórias (Simar, 2000). Nesse sentido, o uso do procedimento bootstrap torna-se uma ferramenta indispensável para conferir maior robustez estatística às estimativas de eficiência.

O método bootstrap, introduzido na literatura de DEA por Simar e Wilson (1998, 2007), consiste na geração de réplicas da amostra original por meio de reamostragens com reposição. Cria um grande número de novas fronteiras artificiais baseadas em variações dos seus dados originais, permitindo estimar a distribuição da eficiência de cada DMU, construir intervalos de confiança e corrigir o viés dos escores de eficiência. Assim, além da estimativa pontual da eficiência técnica, é possível avaliar a sua confiabilidade estatística e obter medidas ajustadas que se aproximam mais da verdadeira eficiência populacional.

Matematicamente, o procedimento pode ser descrito em três etapas principais:

1. Estima-se a eficiência técnica inicial das DMUs pelo modelo DEA (CCR ou BCC) $\hat{\varphi}_i$;
2. A partir dos resíduos amostrais (ou seja, do conjunto de todas as variações possíveis que o escore de eficiência de uma DMU pode assumir devido a pequenas mudanças na amostra), gera-se uma distribuição empírica por meio de reamostragem bootstrap;
3. Calcula-se a eficiência ajustada (bias-corrected efficiency), bem como intervalos de confiança para os escores originais.

O viés da eficiência inicial pode ser expressa como:

$$\widehat{viés}_i = \frac{1}{B} \sum_{b=1}^B \hat{\varphi}_i^* - \hat{\varphi}_i = \bar{\varphi}_i^* - \hat{\varphi}_i \quad (14)$$

Nessa expressão, $\bar{\varphi}_i^*$ é a média das B estimativas de eficiência bootstrap da DMUi. Daí, com o viés estimado, o estimador corrigido $\tilde{\varphi}_i$ é:

$$\tilde{\varphi}_i = \hat{\varphi}_i - \widehat{viés}_i = \hat{\varphi}_i - \bar{\varphi}_i^* + \hat{\varphi}_i = 2\hat{\varphi}_i - \bar{\varphi}_i^* \quad (15)$$

Observa-se também que, conhecendo a distribuição empírica da eficiência bootstrap e do viés, é possível realizar testes de hipóteses e encontrar intervalos de confiança com um nível de significância $(1-\alpha)$, calculando os valores (quartis) críticos a_α e b_α , de forma que:

$$P[-b_\alpha \leq \tilde{\varphi}_i \leq -a_\alpha] \approx P[-b_\alpha \leq \tilde{\varphi}_i \leq -a_\alpha] \approx 1 - \alpha \quad (16)$$

Nota-se que o algoritmo de bootstrap exige a execução de diversos passos e uma demanda computacional relativamente intensa. Para operacionalizar esse procedimento, foi desenvolvida uma função auxiliar em R, implementada com o pacote Benchmarking, que sistematiza todas as etapas necessárias. Essa função possibilita estimar uma fronteira não determinística (DEA VRS orientado a outputs), dando propriedades estatísticas ao DEA, corrigindo o viés dos escores de eficiência e construindo intervalos de confiança, permitindo maior transparência e reprodutibilidade na análise.

No campo da educação superior, a utilização do bootstrap em conjunto com a DEA já se consolidou como prática metodológica em estudos nacionais e internacionais (Arbula Blecich, 2024; Santos; Silva Júnior; Nunes, 2022). Tal abordagem permite não apenas maior precisão nos resultados, mas também a realização de inferências estatísticas, fortalecendo a validade das análises.

No presente estudo, o procedimento bootstrap foi implementado de forma manual no RStudio, com base no pacote Benchmarking, totalizando 500 reamostragens. Esse quantitativo foi definido de acordo com recomendações da literatura, de modo a equilibrar robustez estatística e viabilidade computacional. Ressalta-se que o uso do bootstrap mostra-se particularmente relevante no contexto investigado, dada a heterogeneidade entre IES públicas e privadas e a diversidade estrutural das instituições vinculadas ao Fies, o que reforça a necessidade de procedimentos que assegurem estimativas estáveis e confiáveis.

3.7 ANÁLISE DO IMPACTO DAS VARIÁVEIS AMBIENTAIS

A DEA é amplamente utilizada para mensurar a eficiência relativa entre DMUs. No entanto, como destacam Fried, Lovell e Schmidt (2008), essa técnica não incorpora de forma explícita fatores externos ao processo produtivo que podem afetar os resultados obtidos. Tais fatores — denominados variáveis ambientais ou contextuais — referem-se a elementos institucionais, sociais e econômicos que não

estão diretamente sob o controle da gestão das DMUs, mas que exercem influência relevante sobre seus níveis de eficiência.

Para lidar com essa limitação, diversos autores recomendam o emprego de uma abordagem em dois estágios. No primeiro estágio, estima-se a eficiência técnica por meio da DEA. No segundo estágio, os escores de eficiência obtidos são utilizados como variáveis dependentes em modelos estatísticos que permitem avaliar a influência das variáveis ambientais (Ray, 2004; Hoff, 2007).

Um aspecto metodológico central reside na escolha do modelo de regressão. Por se tratar de índices censurados no intervalo $[0,1]$ que não seguem uma distribuição normal, a literatura tem apontado limitações no uso da regressão linear tradicional (OLS). Nesse sentido, Simar e Wilson (2007) sugeriram o emprego do modelo Tobit com correção via *bootstrap*, o qual se consolidou como uma prática recorrente em estudos aplicados, dada sua capacidade de lidar com o viés e assegurar maior robustez estatística (McDonald, 2009).

Contudo, no presente estudo, optou-se por empregar técnicas de *Machine Learning* (ML) na análise do impacto das variáveis ambientais sobre a eficiência. Essa escolha decorre da necessidade de maior flexibilidade analítica e da adequação dessas técnicas às características da base de dados, marcada por heterogeneidade institucional e possíveis relações não lineares entre as variáveis. Assim, embora o modelo Tobit com correção via *bootstrap* constitua uma abordagem consolidada na literatura, a utilização de ML permite investigar de forma mais abrangente e consistente a influência de fatores contextuais sobre a eficiência das IES.

A operacionalização dessa estratégia metodológica baseou-se na seleção de algoritmos de ML segundo três critérios centrais: a) adequação estatística à natureza dos dados; b) grau de interpretabilidade dos resultados; e c) capacidade de capturar padrões não lineares e interações complexas entre as variáveis ambientais. Com base nesses critérios, foram empregados os seguintes modelos:

O Elastic Net (Zou; Hastie, 2005) combina as penalizações L1 (LASSO) e L2 (Ridge), permitindo simultaneamente a seleção de variáveis relevantes e a mitigação de problemas de multicolinearidade. Essa característica torna o modelo particularmente adequado para contextos com múltiplas variáveis correlacionadas, além de aproximá-lo do paradigma econométrico tradicional, ao produzir uma relação funcional explícita e interpretável entre eficiência e fatores contextuais (Kuhn; Johnson, 2013).

O Random Forest (Breiman, 2001), por sua vez, consiste em um método de ensemble baseado em árvores de decisão, robusto ao tratamento conjunto de variáveis contínuas e categóricas. Sua principal vantagem reside na capacidade de capturar interações não especificadas *a priori* e relações não lineares, além de fornecer métricas de importância das variáveis, ampliando a compreensão dos fatores que mais influenciam os níveis de eficiência observados.

Complementarmente, foi empregado o XGBoost (*Extreme Gradient Boosting*), proposto por Chen e Guestrin (2016), reconhecido na literatura recente pela elevada acurácia preditiva. Trata-se de um algoritmo baseado em gradient boosting que combina eficiência computacional com grande flexibilidade funcional, sendo especialmente eficaz em cenários caracterizados por elevada complexidade estrutural e múltiplas interações entre variáveis explicativas.

A aplicação conjunta dessas técnicas permite equilibrar compreensibilidade e capacidade preditiva. Enquanto o Elastic Net oferece maior proximidade com abordagens econométricas tradicionais, favorecendo a interpretação dos coeficientes estimados, o Random Forest e o XGBoost ampliam o poder explicativo do modelo ao explorar relações não lineares e padrões complexos nos dados. Evidências recentes indicam que a combinação de modelos lineares regularizados com algoritmos de ensemble resulta em análises mais robustas e consistentes em estudos aplicados à eficiência educacional (Chen *et al.*, 2025; Arbula Blecich, 2024).

A estimação dos modelos de aprendizado de máquina foi precedida por um processo sistemático de calibração de hiperparâmetros (*tuning*), com o objetivo de maximizar a capacidade de generalização dos algoritmos e reduzir riscos de sobreajuste. Para esse fim, adotou-se uma estratégia de busca em grades (*grid search*), combinada com validação cruzada, permitindo avaliar o desempenho dos modelos sob diferentes configurações paramétricas e selecionar aquelas que apresentaram melhor equilíbrio entre viés e variância.

No caso do Elastic Net, foram definidos grids em escala logarítmica para o parâmetro de penalização (*penalty*), variando entre 10^{-4} e 10^1 , bem como valores intermediários para o parâmetro de mistura (*mixture*), responsável por controlar a ponderação relativa entre as penalizações LASSO (L1) e Ridge (L2). Essa especificação buscou evitar soluções concentradas nos extremos do espaço paramétrico, ampliando a possibilidade de obtenção de modelos parcimoniosos,

porém com adequada capacidade explicativa, especialmente em contextos caracterizados por multicolinearidade entre variáveis ambientais.

Para o Random Forest, o processo de calibração envolveu o ajuste do número de variáveis candidatas por divisão (*mtry*) e do tamanho mínimo dos nós terminais (*min_n*). Adicionalmente, o número de árvores foi ampliado de 1.000 para 2.000, com o propósito de reduzir a variância do estimador e aumentar a estabilidade preditiva do modelo. Essa estratégia é consistente com a literatura sobre métodos de ensemble baseados em árvores, que recomenda o uso de florestas mais extensas para mitigar flutuações amostrais e melhorar a robustez dos resultados.

No XGBoost, a calibração concentrou-se na ampliação do espaço de busca dos principais hiperparâmetros estruturais, incluindo a profundidade das árvores (*tree_depth*, variando de 3 a 8), a taxa de aprendizado (*learn_rate*, entre 0,001 e 0,1) e o parâmetro de regularização associado à redução de perda (*gamma*). Além disso, foram consideradas variações explícitas nos parâmetros de amostragem de variáveis (*colsample_bytree*) e de observações (*sample_size*). Esses ajustes visaram equilibrar a capacidade de generalização do algoritmo, reduzir riscos de sobreajuste e permitir a captura de padrões não lineares mais sutis presentes nos dados.

De forma geral, o procedimento de calibração adotado assegura que as diferenças de desempenho observadas entre os algoritmos não decorram de escolhas arbitrárias de parametrização, mas reflitam características estruturais dos próprios modelos. Essa etapa reforça a consistência metodológica da análise e contribui para a robustez das inferências sobre o impacto das variáveis ambientais na eficiência das Instituições de Ensino Superior.

Dessa forma, a estratégia metodológica adotada — que combina a aplicação de diferentes algoritmos de aprendizado de máquina com procedimentos sistemáticos de calibração de parâmetros — mostra-se adequada às características da base de dados e aos objetivos do estudo, ao possibilitar uma análise abrangente e consistente do impacto das variáveis ambientais sobre a eficiência das Instituições de Ensino Superior (IES). Em consonância com essa abordagem, foram selecionadas variáveis ambientais capazes de captar dimensões socioeconômicas, institucionais e regionais que, embora não estejam sob controle direto da gestão, exercem influência relevante sobre o desempenho observado das instituições.

Especificamente, foram consideradas as seguintes variáveis ambientais:

- Produto Interno Bruto per capita municipal (PIBPC_MUN): aproxima o nível de desenvolvimento socioeconômico do entorno imediato da IES.
- Densidade demográfica municipal (DENS_MUN): indica a concentração populacional, refletindo potenciais pressões sobre a infraestrutura e serviços educacionais.
- Tipo de rede (TP_REDE): variável dummy que assume valor 0 para instituições públicas e 1 para instituições privadas, captando diferenças relacionadas a governança, fontes de financiamento e mecanismos de accountability.
- Idade institucional (ANO_MEC): calculada a partir do ano de credenciamento da IES junto ao MEC, representa o grau de consolidação acadêmica e administrativa.
- Sede na capital (IN_CAPITAL): variável dummy que assume valor 0 para instituições sediadas em municípios do interior e 1 para instituições sediadas em capitais estaduais, destinada a captar possíveis vantagens locais, como melhor infraestrutura e maior proximidade com centros decisórios.
- Latitude e Longitude (LAT/LON): fornecem a localização geográfica exata da sede da IES, permitindo análises espaciais, detecção de clusters e exploração de padrões territoriais de eficiência.
- Participação no Fies (DUMMY_FIES): dummy (0/1) que assume 1 quando a IES possui ao menos uma matrícula financiada pelo Fies no ano t ; 0, caso contrário.

O Quadro 4 sintetiza a descrição dessas variáveis, a metodologia adotada para sua construção e os principais estudos que respaldam sua utilização na literatura.

Quadro 4 - Variáveis Ambientais

Variável	Descrição	Metodologia	Estudos Anteriores
PIB_TOTAL	Produto Interno Bruto Total	PIB Total correspondente ao total dos PIB de cada município, segundo dados do IBGE.	Estudos como Arbula Bleich (2024) destacam o PIB como indicador-chave na eficiência educacional, refletindo a capacidade econômica das regiões analisadas.
DENS	Densidade Demográfica	Densidade Demográfica de cada município, considerando as populações e dimensões territoriais de cada município que as compõe.	Afonso <i>et al.</i> (2005) identificam que a densidade populacional influencia a alocação de

TP_REDE	Rede de Ensino	Tipo de Rede de Ensino da IES: 1 – Pública; 2 – Privada.	recursos e a capacidade operacional das IES. Santos <i>et al.</i> (2022) apontam que IES privadas são mais eficientes devido à maior flexibilidade administrativa e foco em resultados financeiros.
IDADE	Ano de Credenciamento	Ano de credenciamento da IES junto ao MEC como forma de mensurar o tempo de existência da instituição.	Siqueira (2015) sugere que universidades mais consolidadas, que tiveram tempo para expandir suas operações e aprimorar seus processos, tendem a apresentar maior eficiência operacional.
IN_CAPITAL	Sede na Capital	Identifica se a sede da IES se encontra na capital do estado.	Pereira (2023) sugere que IES sediadas em capitais tendem a ser mais eficientes.
LAT e LON	Latitude e Longitude	A localização geográfica é um fator relacionado com a eficiência e a produtividade.	Pereira (2023) conclui que a localização influencia na eficiência das Instituições de Ensino Superior.
DUMMY_FIES	Indicador de participação da IES no Fies.	Variável dummy binária (0/1) que identifica se a instituição possui matrículas financiadas pelo Fies no ano t ; assume valor 1 quando há ao menos uma matrícula financiada e 0, caso contrário. Fonte: Censo da Educação Superior/Inep e FNDE.	Souza (2021); Pontuschka (2016); Becker e Mendonça (2021) destacam os efeitos do Fies sobre acesso, permanência e eficiência institucional.

Fonte: (Brasil, [s.d.].c). Dicionário de Dados.

A literatura empírica evidencia que variáveis ambientais exercem impactos diferenciados sobre a eficiência das instituições de ensino. Nazarko e Šaparauskas (2014) ressaltam que assimetrias regionais afetam de maneira significativa tanto a alocação de recursos quanto a qualidade do ensino ofertado. No contexto brasileiro, Santos, Silva Júnior e Nunes (2022) demonstram que fatores estruturais, como o tempo de credenciamento e o porte institucional, estão associados a desempenhos médios superiores, reforçando a necessidade de considerar tais condicionantes na avaliação da eficiência das IES.

Dessa forma, a consideração das variáveis ambientais torna-se essencial para evitar interpretações enviesadas dos escores de eficiência estimados. A distinção entre fatores passíveis de controle pela gestão interna das instituições e aqueles decorrentes de restrições externas permite não apenas uma leitura mais precisa das fontes de ineficiência, mas também a formulação de recomendações mais

consistentes para gestores institucionais e para o desenho de políticas públicas no âmbito do ensino superior.

Complementarmente à análise multivariada conduzida por meio de técnicas de aprendizado de máquina, faz-se necessário aprofundar a investigação sobre um dos principais elementos de política pública associados ao financiamento do ensino superior privado no Brasil: o Fies. Nesse sentido, a seção seguinte dedica-se a uma etapa explicativa adicional, de natureza econométrica, voltada à avaliação estatística da associação entre a participação no Fies e os níveis de eficiência operacional das Instituições de Ensino Superior privadas.

3.8 ANÁLISE EXPLICATIVA DOS EFEITOS DO FIES POR BETA REGRESSION

Após a mensuração dos níveis de eficiência e a análise das variáveis ambientais por meio de algoritmos de aprendizado de máquina, realizou-se uma etapa explicativa complementar com o objetivo de verificar estatisticamente se a participação no Programa de Financiamento Estudantil (Fies) está associada a maiores níveis de eficiência operacional entre as Instituições de Ensino Superior privadas.

Para isso, foi utilizado o modelo de regressão Beta (Beta Regression), apropriado para variáveis dependentes contínuas limitadas no intervalo (0,1) e que não tenham distribuição normal, como é o caso dos escores de eficiência global. O modelo permite estimar o efeito marginal das variáveis independentes sobre a eficiência e testar a significância estatística de cada coeficiente. Ela foi proposta por Ferrari e Cribari-Neto (2004). Relaciona as variáveis explicativas (ambientais ou exógenas) com o valor esperado da eficiência, usando uma "função de ligação" (geralmente a logit) para garantir que o resultado fique sempre entre 0 e 1. A regressão Beta é muito mais rigorosa que o Tobit para lidar com a natureza "achatada" e limitada dos índices de eficiência, resultando em coeficientes mais confiáveis.

A variável dependente utilizada foi a eficiência global (eg), e as variáveis explicativas incluíram indicadores contextuais (densidade demográfica, PIB per capita municipal, idade institucional, localização geográfica, tipo de rede e sede na capital), além da variável `dummy_Fies`, que assume valor 1 para IES com participação no programa e 0, caso contrário.

O modelo foi estimado utilizando o pacote `betareg` no RStudio, e os resultados foram complementados por um teste t de diferença de médias, comparando o

desempenho médio das IES participantes e não participantes do Fies. Essas análises permitem identificar não apenas a relevância preditiva observada na etapa de aprendizado de máquina, mas também a significância estatística da associação entre o Fies e os níveis de eficiência das IES privadas.

3.9 AVALIAÇÃO DINÂMICA E ÍNDICE DE PRODUTIVIDADE DE MALMQUIST

A análise de eficiência tradicional, baseada em modelos DEA estáticos, avalia a relação entre insumos e produtos em um único período. Entretanto, as IES operam em contextos dinâmicos, sujeitos a mudanças tecnológicas, políticas educacionais e transformações socioeconômicas. Nesses cenários, torna-se necessário adotar medidas capazes de capturar a evolução temporal da produtividade.

Nesse sentido, o Índice de Produtividade de Malmquist (IM) é amplamente utilizado para decompor variações de desempenho ao longo do tempo. Introduzido originalmente por Malmquist (1953) em estudos sobre teoria do consumo, o índice foi posteriormente adaptado por Caves, Christensen e Diewert (1982) e operacionalizado em DEA por Färe, Grosskopf, Norris e Zhang (1994). Sua formulação permite comparar a produtividade de uma unidade em dois períodos consecutivos, distinguindo entre ganhos decorrentes de mudanças na eficiência relativa e aqueles associados a mudanças tecnológicas.

Matematicamente, o índice de Malmquist é definido a partir das funções distância orientadas a outputs (inversas do índice de eficiência de Farrell), e pode ser expresso da seguinte forma:

$$M_t^{t+1}(x_t, y_t, x_{t+1}, y_{t+1}) = \left[\frac{D_t^0(x_{t+1}, y_{t+1})}{D_t^0(x_t, y_t)} \cdot \frac{D_{t+1}^0(x_{t+1}, y_{t+1})}{D_{t+1}^0(x_t, y_t)} \right]^{1/2} \quad (17)$$

Onde:

- x_t, y_t representam, respectivamente, insumos e produtos no período t ;
- $D_t^0(x, y)$ é a função distância orientada a outputs, calculada em relação à tecnologia de referência no período t .

O índice M_t^{t+1} pode ser decomposto em duas componentes fundamentais:

- Mudança na eficiência técnica (catch-up effect): reflete a capacidade da DMU em se aproximar da fronteira eficiente ao longo do tempo, representando ganhos de gestão, reorganização interna ou melhor aproveitamento dos recursos disponíveis.

- Mudança tecnológica (frontier-shift effect): capta o deslocamento da própria fronteira de produção, indicando progresso tecnológico, inovação ou transformações estruturais no setor.

Dessa forma, valores de $M > 1$ indicam crescimento da produtividade total dos fatores, enquanto valores de $M < 1$ apontam retração.

A literatura aplicada ao setor educacional reforça a utilidade desse índice. No Brasil, Siqueira (2015) identificou ganhos modestos de produtividade nas IESs federais, associados principalmente ao aumento de escala após a implantação do REUNI. Santos, Silva Júnior e Nunes (2022) observaram que o efeito catch-up foi mais relevante que o frontier-shift, sugerindo que as melhorias decorreram de maior eficiência relativa das instituições, mais do que de inovações tecnológicas sistêmicas. Em âmbito internacional, Arbula Blecich (2024) destacou a pertinência da abordagem dinâmica para captar variações significativas de eficiência no ensino superior europeu.

No contexto desta dissertação, a aplicação do índice de Malmquist permitirá avaliar não apenas se as IES públicas federais e as IES privadas foram eficientes em cada período, mas também como a produtividade evoluiu entre 2019 e 2023. Isso possibilitará distinguir os ganhos associados a melhorias gerenciais internas (eficiência técnica) daqueles decorrentes de transformações estruturais no sistema educacional (mudança tecnológica).

Essa formulação teórica fornece a base metodológica para a etapa seguinte, em que serão examinados os índices obtidos a fim de identificar se a produtividade do ensino superior brasileiro evoluiu predominantemente em função de ganhos de eficiência interna ou de mudanças tecnológicas mais amplas no setor.

4 ANÁLISE DOS RESULTADOS

Este capítulo reúne e discute os resultados obtidos a partir da aplicação da DEA e de seus desdobramentos no contexto das IES públicas federais e privadas brasileiras, com destaque para aquelas vinculadas ao Fies.

A análise inicia-se pela apresentação descritiva das variáveis utilizadas, permitindo identificar padrões e variações estruturais no período de 2019 a 2023. Em seguida, procede-se à identificação dos outliers, destacando os casos de supereficiência, que auxiliam na compreensão de situações extremas e de sua influência sobre a fronteira de eficiência.

Inicialmente, procedeu-se ao cálculo dos índices de eficiência, os quais servem de base para as análises subsequentes. A partir desses resultados, realiza-se o teste de modelo, com o objetivo de definir a adequação entre retornos constantes ou variáveis de escala. Em seguida, desenvolve-se a análise comparativa entre grupos, voltada à identificação de diferenças de desempenho entre as IES públicas federais e privadas. Posteriormente, são apresentados os resultados do modelo em rede (NDEA), tanto em sua forma clássica quanto na versão corrigida pelo bootstrap orientado a outputs, destacando-se os procedimentos de cálculo, a correção de viés e a interpretação dos resultados associados aos gastos diretos e indiretos com o ensino superior.

O capítulo prossegue com a análise do impacto das variáveis ambientais, que busca identificar fatores exógenos capazes de influenciar a eficiência, distinguindo aqueles sob controle da gestão interna das IES daqueles que refletem restrições contextuais.

Por fim, aborda-se a avaliação dinâmica da produtividade com base no Índice de Malmquist, o que permite verificar não apenas o desempenho em cada período, mas também a evolução da produtividade ao longo dos cinco anos analisados, diferenciando os efeitos derivados de ganhos de eficiência técnica daqueles associados a transformações tecnológicas mais amplas no sistema educacional.

4.1 ANÁLISE DESCRITIVA

Em um primeiro passo, procedeu-se à estatística descritiva sobre as IES analisadas. Para assegurar a consistência do painel de dados, foi realizado o balanceamento da amostra, restringindo a base às IES que apresentaram registros

completos em todos os anos considerados no recorte temporal de 2019 a 2023. Esse tratamento permitiu eliminar instituições com informações ausentes em algum dos períodos, garantindo que a comparação intertemporal fosse realizada sobre o mesmo conjunto de unidades. Como resultado, obteve-se um painel balanceado com 1.885 IES, conforme detalhado na Tabela 1, o que preserva a homogeneidade da amostra e evita vieses decorrentes de flutuações na presença das instituições ao longo do tempo.

Tabela 1 - Caracterização das IES da amostra (2019–2023)

Atributo / Indicador	Total (N=1.885)	Públicas Federais	Privadas	Observações
Número de IES	1.885	109	1.776	Painel balanceado (2019–2023)
Distribuição geográfica (%)				Proporção por macrorregião
– Norte	10.72 %	15.6 %	7.5 %	Menor participação privada
– Nordeste	22.86 %	25.7 %	22.7 %	Segunda maior concentração
– Sudeste	41.91 %	33.0 %	42.4 %	Maior concentração
– Sul	16.50 %	15.6 %	16,6 %	Terceira maior concentração
– Centro-Oeste	10.72 %	10.1 %	10.8 %	Distribuição pública x privada
Média de matrículas anuais	8.256.261 4.380	1.306.540 11.986	6.949.721 3.913	Valores absolutos Relativo por IES
Média de docentes	299.987 159	118.438 1.086	181.549 102	Inclui regime integral e parcial
Média de técnicos	298.078 158	128.107 1.175	169.971 96	Medida da infraestrutura de apoio
Idade média das instituições (anos)	26,3	45,6	25,2	Calculada a partir do credenciamento MEC
Sede em capitais (%)	36,4 %	59,6 %	35,0 %	Proxy para desigualdade urbano-interior

Fonte: Elaboração própria a partir dos microdados do Censo da Educação Superior (Brasil, [s.d.]).

No que diz respeito às matrículas, observa-se que as IES privadas concentram a maior parte dos estudantes de ensino superior no Brasil, evidenciando seu papel central no processo de expansão do acesso, especialmente no contexto de políticas públicas de financiamento estudantil, como o Fies. Por sua vez, as IES públicas federais, embora representem uma parcela menor do total de matrículas, apresentam características institucionais distintas, associadas a uma trajetória de expansão vinculada a programas governamentais específicos. Essa configuração é consistente com a análise de Siqueira (2015), segundo a qual a ampliação do sistema federal de ensino superior ocorreu de forma planejada e induzida por políticas públicas, enquanto a iniciativa privada respondeu de maneira mais imediata ao crescimento da demanda. Assim, a comparação entre os dois segmentos evidencia estratégias diferenciadas de

atuação no sistema de ensino superior brasileiro, aspecto relevante para a análise subsequente de eficiência.

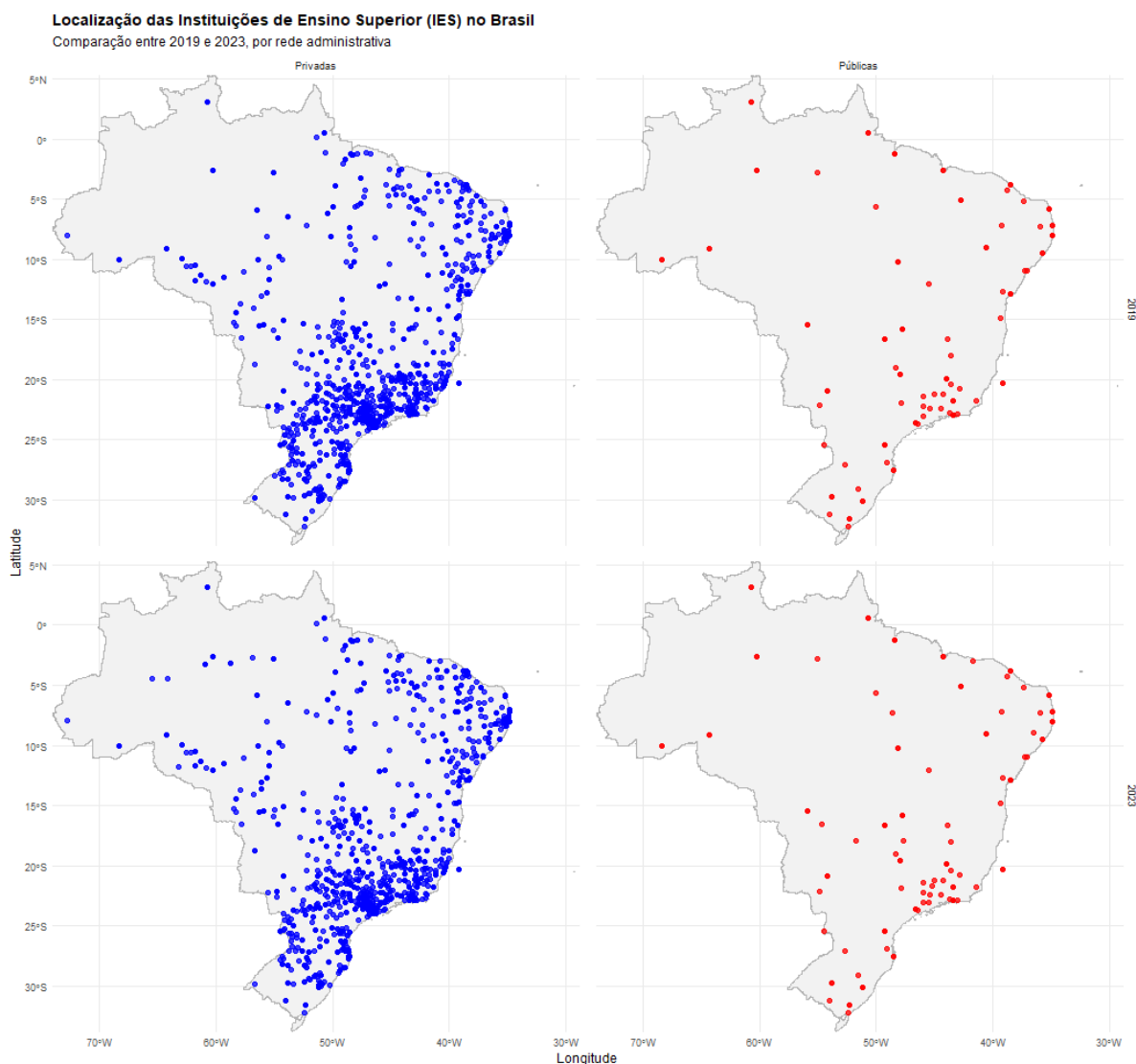
No que se refere ao quadro de pessoal, observa-se uma diferença estrutural relevante entre as IES públicas federais e as privadas. As IES federais contam, em média, com número mais elevado de docentes e técnicos administrativos, reflexo de sua maior dimensão organizacional e da obrigatoriedade legal de manter corpo docente com elevada titulação acadêmica (Siqueira, 2015). Já entre as instituições privadas, verifica-se maior heterogeneidade: enquanto parte delas apresenta estruturas enxutas, com menor número de docentes em regime integral e de técnicos de apoio, outras, especialmente as de grande porte, aproximam-se do padrão das públicas. Essa disparidade corrobora os achados de Santos, Silva Júnior e Nunes (2022), segundo os quais fatores institucionais e de governança impactam diretamente na eficiência relativa das IES. Assim, o comparativo evidencia que, apesar de possuírem mais recursos humanos, as federais enfrentam desafios relacionados à rigidez administrativa, enquanto as privadas, ainda que mais flexíveis, apresentam limitações na manutenção de equipes amplas e qualificadas.

Em relação à idade das instituições, observa-se que as IES públicas federais concentram as unidades mais antigas do país, com diversas universidades fundadas ainda no século XX e consolidadas como referência nacional. Já entre as instituições privadas, predominam aquelas criadas a partir das décadas de 1980 e 1990, no contexto de expansão da educação superior. A média de tempo de existência da amostra analisada é de aproximadamente 27 anos, indicando que, de forma geral, as instituições já percorreram trajetória significativa de consolidação. Esse resultado sugere que, no período de 2019 a 2023, a maior parte das IES ativas no painel já havia ultrapassado sua fase inicial de estruturação, o que confere maior estabilidade à análise de eficiência realizada.

A distribuição geográfica das IES revelou uma concentração significativa na Região Sudeste, na média dos anos da série, seguida atualmente pela Região Nordeste, passando a Região Sul, vindo a representar parcela cada vez mais relevante do total de instituições. Em contrapartida, as Regiões Centro-Oeste e Norte não chegaram a atingir 16% do total em nenhum dos anos analisados. Esse quadro evidencia a desigualdade regional na oferta de ensino superior, refletindo o predomínio histórico das regiões mais desenvolvidas economicamente na instalação e expansão das IES. Para ilustrar esse movimento, a Figura 6 apresenta a distribuição

espacial das Instituições de Ensino Superior (IES) no Brasil, segmentadas por rede administrativa (privada e pública), considerando os anos de 2019 e 2023.

Figura 6 - Distribuição das IES do Brasil (2019 e 2023)



Fonte: Elaboração própria a partir dos microdados do Censo da Educação Superior (Brasil, [s.d.]).

A visualização cartográfica evidencia a expressiva predominância das IES privadas, caracterizadas por elevada densidade espacial, especialmente nas regiões Sudeste e Sul, bem como ao longo do litoral brasileiro, refletindo a maior concentração populacional e econômica do país. As IES públicas, por sua vez, aparecem em número significativamente menor e com distribuição mais dispersa, associada à lógica de cobertura territorial e interiorização do ensino superior.

Do ponto de vista quantitativo, observa-se que o número total de IES reduziu levemente de 2.416 em 2019 para 2.385 em 2023. Essa variação decorre, sobretudo, da dinâmica das instituições privadas, cujo quantitativo passou de 2.306 IES em 2019

para 2.264 em 2023, indicando um processo recente de ajuste e estabilização no setor. Em contraste, o número de IES públicas apresentou crescimento no período, passando de 110 instituições em 2019 para 121 em 2023, ainda que tal expansão seja numericamente modesta quando comparada ao universo total de instituições.

Em termos espaciais, a comparação entre 2019 e 2023 indica que, apesar das variações no número total de instituições, a configuração geográfica do sistema de ensino superior brasileiro permanece relativamente estável, mantendo-se a concentração das IES privadas nos principais centros urbanos e a presença estratégica das IES públicas em diferentes regiões do território nacional.

Tais resultados reforçam a importância de considerar as variáveis contextuais e regionais na análise da eficiência, uma vez que as condições socioeconômicas locais podem influenciar tanto a disponibilidade de recursos quanto a demanda por serviços educacionais. Além disso, a comparação entre IES públicas federais e privadas ganha relevância nesse cenário, pois tais diferenças estruturais podem ser intensificadas pela desigualdade geográfica.

Uma das justificativas para a maior concentração de instituições na região Sudeste pode ser atribuída à sua elevada densidade demográfica, fator já apontado por Afonso *et al.* (2005) como determinante na alocação de recursos e na capacidade operacional das IES. Todavia, esse argumento não se aplica de forma uniforme às demais regiões, uma vez que a região Nordeste, apesar de apresentar população significativamente superior à da região Sul, ainda mantém proporção menor de instituições.

Outro aspecto relevante é o nível de renda regional. Arbula Blecich (2024) destaca que o Produto Interno Bruto (PIB) é um indicador-chave para a eficiência educacional, refletindo a capacidade econômica de cada região em sustentar maior número de instituições. Da mesma forma, Santos, Silva Júnior e Nunes (2022) ressaltam que fatores estruturais, como porte e financiamento institucional, influenciam diretamente os resultados de eficiência. Assim, a concentração de empresas, indústrias e centros de serviços nas regiões Sul e Sudeste, somada ao maior dinamismo econômico, tende a gerar demanda mais elevada por profissionais qualificados em áreas como Administração e Contabilidade, estimulando a instalação de novas IES, sobretudo privadas.

Além disso, Pereira (2023) aponta que a localização geográfica exerce papel central na eficiência, com instituições situadas em capitais apresentando desempenho

diferenciado em relação às aquelas do interior. Essa evidência ajuda a explicar por que, nas regiões Norte, Nordeste e Centro-Oeste, as IES permanecem concentradas em capitais ou áreas metropolitanas, enquanto no Sul e Sudeste observa-se uma distribuição mais homogênea pelo interior dos estados.

Esse conjunto de fatores reforça que a desigualdade regional observada na Tabela 1 e na Figura 6 não se limita a diferenças populacionais, mas resulta da interação entre variáveis socioeconômicas, institucionais e geográficas que moldam a oferta de ensino superior no Brasil.

Dando continuidade à análise, o próximo passo consiste na apresentação dos dados utilizados para o cálculo da DEA. Para essa etapa, elaborou-se uma tabela estatística contemplando as principais variáveis discricionárias que compõem o primeiro estágio do modelo em rede (NDEA). A construção desse material foi realizada com o auxílio do software R, a partir dos microdados do Censo da Educação Superior disponibilizados pelo Inep.

A Tabela 2 apresenta as estatísticas descritivas dos insumos e produtos em sua forma original, evidenciando elevada dispersão entre as instituições, especialmente no número de matrículas e no somatório de acervo bibliográfico digital. Esse comportamento reflete a heterogeneidade estrutural das IES públicas federais e privadas, com diferenças marcantes no porte e na capacidade produtiva.

Tabela 2 - Análise descritiva dos insumos e produtos – originais

Variável	Mínimo	Q1	Mediana	Média	Q3	Máximo
QT_MAT	0,001	216	740	4.379,98	2.494	771.858
QT_DOC_EXE	0,001	23	51	159,14	127	5.812
QT_CONC	0,001	31	125	628,25	403	123.474
SOMATORIO_SERVICOS	0,001	4	5	4,89	6	7
SOMATORIO_ACERVO	0,001	1.840	12.000	45.315,39	29.400	13.008.000
QT_ABANDONO	0,001	33	135	1.298,46	529	413.642

Fonte: Elaboração própria com base nos dados do Inep (Brasil, [s.d.]c).

Com o objetivo de tornar os dados mais comparáveis e reduzir os efeitos da escala, as variáveis foram normalizadas e tratadas conforme metodologia descrita anteriormente. Os resultados desse processo estão sintetizados na Tabela 3, que apresenta as mesmas estatísticas descritivas já ajustadas. A transformação evidencia maior homogeneidade nos valores, preservando, contudo, as diferenças relativas entre as instituições.

Tabela 3 - Análise descritiva dos insumos e produtos – tratados

Variável	Mínimo	Q1	Mediana	Média	Q3	Máximo
QT_MAT	0,127058	0,405455	0,49795	0,497342	0,591053	0,869451
QT_DOC_EXE	0,05527	0,360664	0,456132	0,469793	0,56426	0,869655
QT_CONC	0,001	0,302778	0,425044	0,401303	0,529112	0,868613
SOMATORIO_SERVICOS	0,531037	0,773867	0,861588	0,83968	0,935754	1
SOMATORIO_PERIODICOS	0,205576	0,481711	0,605644	0,556821	0,665801	0,941936
QT_ABANDONO	0,001	0,28224	0,393518	0,385872	0,503813	0,836171

Fonte: Elaboração própria com base nos dados do Inep (Brasil, [s.d.])c).

Essa estrutura fornece uma base consistente para a análise subsequente, permitindo que os escores de eficiência calculados reflitam de forma mais equilibrada as diferenças reais de desempenho entre as IES analisadas.

A análise descritiva das variáveis (Tabela 3) evidencia diferenças expressivas na distribuição dos indicadores entre as instituições. Observa-se, por exemplo, que o número de matrículas apresenta grande dispersão, variando de apenas uma unidade a mais de 770 mil estudantes. Situação semelhante ocorre no número de docentes em exercício, cujo mínimo registrado é de apenas um professor, ao passo que a instituição de maior porte alcança mais de 5.800 docentes. Esses extremos refletem a heterogeneidade da amostra, que congrega tanto IES federais consolidadas e de grande dimensão quanto instituições privadas de menor porte.

No que se refere aos concluintes, também se observa ampla variação: enquanto algumas instituições registraram valores marginais, outras alcançaram mais de 123 mil formandos no período. Esse contraste se reflete igualmente na produção científica, medida pelo somatório de serviços e publicações periódicas, em que os valores máximos atingem patamares significativamente superior à mediana da amostra. Já o abandono estudantil mostra-se particularmente elevado em determinadas instituições, alcançando mais de 400 mil registros em um único caso, o que aponta para desafios estruturais relacionados à permanência discente.

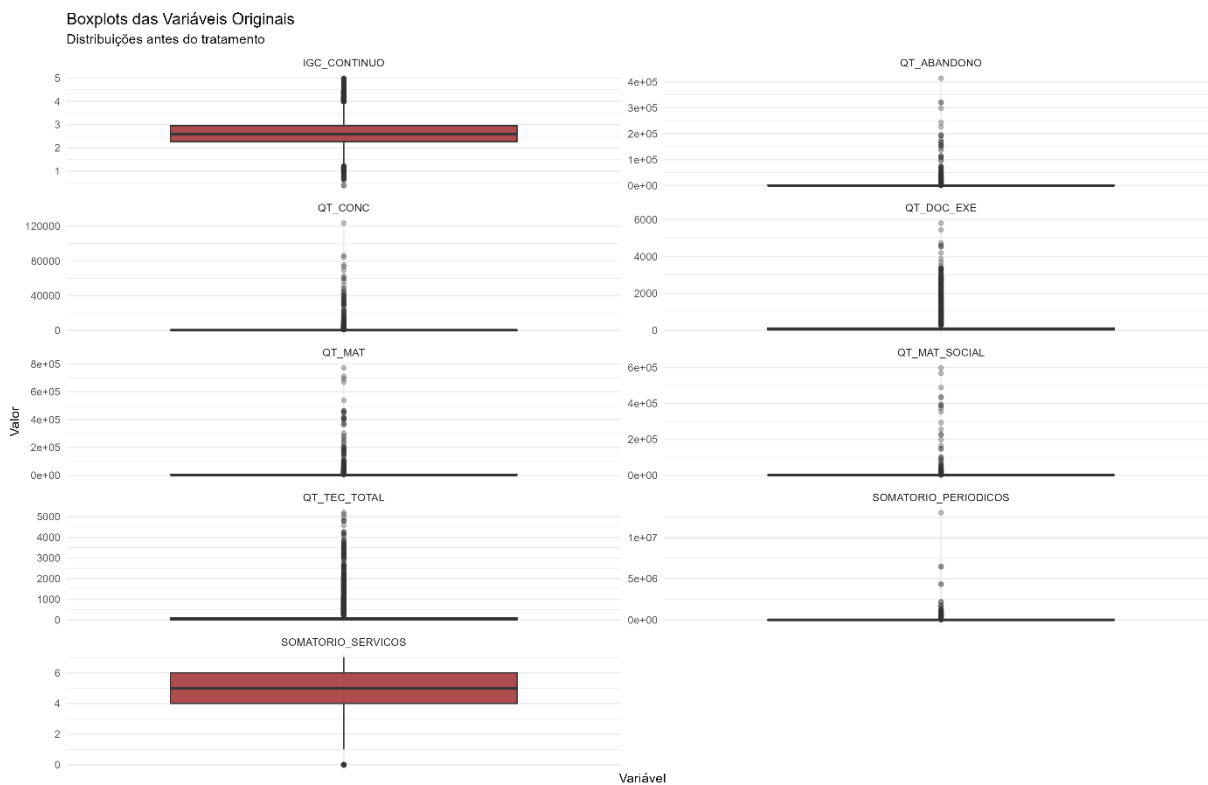
Em síntese, a análise estatística das variáveis confirma a forte heterogeneidade entre as IES públicas federais e privadas, tanto em termos de corpo docente e matrículas quanto na capacidade de conclusão e serviços ofertados. A elevada dispersão de valores reforça a necessidade de uma avaliação robusta, capaz de identificar padrões de eficiência e ineficiência entre instituições de características tão distintas. Nesse sentido, a etapa seguinte volta-se à identificação de outliers, condição essencial para assegurar a consistência da fronteira de eficiência a ser estimada no modelo DEA.

4.2 IDENTIFICAÇÃO DE OUTLIERS

O primeiro passo da análise empírica consistiu na exploração descritiva das variáveis utilizadas no modelo, com o objetivo de compreender sua distribuição e identificar valores extremos sob uma perspectiva univariada. Os boxplots das variáveis originais (Gráfico 1) evidenciam a presença de observações com valores elevados, tanto nos insumos quanto nos produtos, sinalizando elevada heterogeneidade na amostra. Ressalta-se, contudo, que essa análise possui caráter estritamente descritivo, não sendo suficiente, por si só, para a identificação de outliers em um sistema produtivo que envolve múltiplos insumos e produtos, como é o caso da DEA.

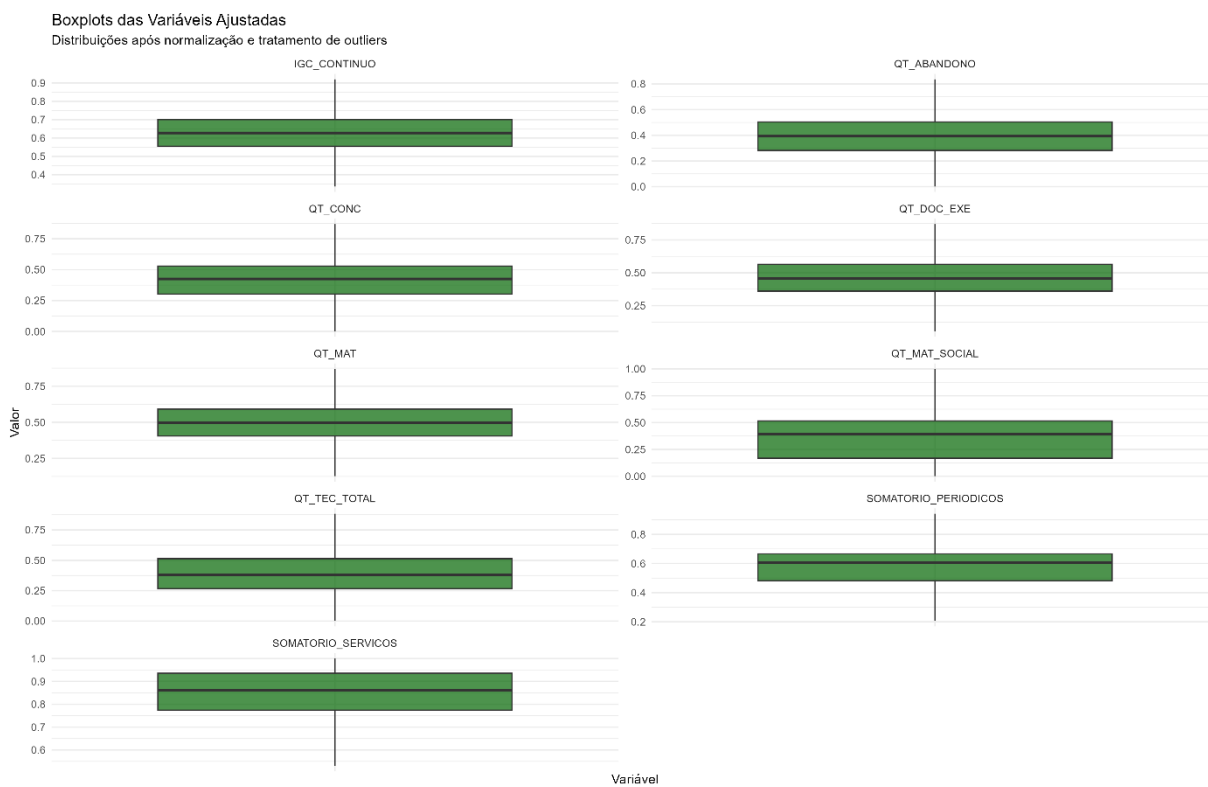
Dessa forma, os procedimentos de normalização e o critério do intervalo interquartil (IQR), apresentados no Gráfico 2, foram empregados exclusivamente como instrumentos auxiliares de padronização e visualização, contribuindo para uma representação mais equilibrada das distribuições e para a comparabilidade preliminar entre as IES. A identificação efetiva de observações atípicas foi realizada posteriormente por meio de abordagens multivariadas compatíveis com a lógica da DEA, baseadas em métodos de fronteira, conforme detalhado na seção metodológica específica, evitando a exclusão indevida de unidades eficientes ou a manutenção de anomalias relevantes sob a ótica tecnológica do sistema analisado.

Gráfico 1 - Boxplots das Variáveis Originais



Fonte: Elaboração própria no R-Studio.

Gráfico 2 - Boxplots das Variáveis Tratadas



Fonte: Elaboração própria no R-Studio.

Antes da realização do cálculo da DEA, procedeu-se à identificação de outliers nas eficiências calculadas, utilizando as técnicas de supereficiência e superineficiência. A aplicação desses métodos permitiu sinalizar instituições com desempenho excepcionalmente elevado, que se posicionam acima da fronteira estimada, bem como aquelas com performance sistematicamente inferior em relação ao conjunto analisado. Ambas as situações, se não tratadas, poderiam distorcer a fronteira de referência e comprometer a robustez das estimativas.

De modo complementar, aplicou-se a técnica da fronteira invertida (Yamada; Matui; Sugiyama, 1994; Entani; Maeda; Tanaka, 2002), na qual inputs e outputs são invertidos para a construção de uma fronteira pessimista, de tal forma que os insumos são considerados produtos e produtos passam a ser insumos. Essa abordagem auxilia na detecção de IES superineficientes, permitindo identificar casos de baixo desempenho consistente, que se afastam de forma relevante do comportamento predominante no conjunto.

4.2.1 Supereficiência

Com base nos cálculos do modelo DEA-VRS orientado a outputs (OO), procedeu-se à análise de supereficiência das IES na Etapa 1 do modelo NDEA, a qual relaciona docentes, técnicos administrativos, serviços e acervo institucional à geração de matrículas. Inicialmente, o modelo DEA clássico permitiu identificar as unidades classificadas como eficientes (DEA = 1). Em seguida, a aplicação do procedimento de supereficiência teve como objetivo avaliar a influência dessas unidades sobre a fronteira eficiente, e não estabelecer um ranking estrito entre DMUs eficientes.

Conforme destacado por Banker e Chang (2006), a supereficiência constitui um instrumento diagnóstico multivariado, capaz de identificar unidades cuja exclusão do conjunto de referência altera de forma relevante a conformação da fronteira DEA. Diferentemente de aplicações em que se observam valores extremos de supereficiência ($SE > 1$), associados a potenciais outliers, os resultados obtidos nesta etapa revelam ausência de escores excepcionalmente elevados, com valores concentrados próximos ou inferiores à unidade. Tal comportamento indica que as DMUs eficientes exercem, em geral, influência moderada e relativamente homogênea sobre a fronteira tecnológica estimada.

Tabela 4 - IES supereficientes Etapa 1 (DEA-VRS OO, 2019–2023)

IES	Nome	Ano	SE_E1
231	FACULDADE DE SÃO BERNARDO DO CAMPO - FASB	2019	0,983422
1205	FACULDADE EDUCACIONAL DA LAPA	2019	0,980088
1679	FACULDADE DE CAMPINA GRANDE DO SUL	2019	0,968566
4185	FACULDADE PERUIBE	2019	0,962401
569	UNIVERSIDADE FEDERAL DO PARÁ	2019	0,903587
1276	INSTITUTO TAUBATÉ DE ENSINO SUPERIOR	2019	0,888778
322	UNIVERSIDADE PAULISTA	2019	0,83804
3375	INSTITUTO DE ENSINO SUPERIOR MÚLTIPLO	2019	0,820702
5669	CENTRO UNIVERSITÁRIO ETEP	2019	0,766674
1263	FACULDADE BEZERRA DE ARAÚJO	2019	0,706718
3625	FACULDADE DO COMPLEXO EDUCACIONAL SANTO ANDRÉ	2019	0,653192
673	CENTRO UNIVERSITÁRIO DA GRANDE DOURADOS	2019	0,638639
527	PONTIFÍCIA UNIVERSIDADE CATÓLICA DE GOIÁS	2019	0,635241
1472	CENTRO UNIVERSITÁRIO LEONARDO DA VINCI	2019	0,443408
802	CENTRO UNIVERSITÁRIO UNICARIOCA	2019	0,395154
2521	INSTITUTO SUPERIOR DE EDUCAÇÃO DO SUL DO PIAUÍ	2019	0,365042
298	UNIVERSIDADE PITÁGORAS UNOPAR	2019	0,349599
580	UNIVERSIDADE FEDERAL DE PERNAMBUCO	2019	0,258417
14069	FACULDADE DE TECNOLOGIA PORTO SUL	2019	0,034483
3777	FACULDADE CURITIBANA	2019	0,003736
1488	FACULDADE LUSÓFONA DO RIO DE JANEIRO	2019	0,003654
1653	FACULDADE DE CIÊNCIA E EDUCAÇÃO DO CAPARAÓ	2020	0,980274
3876	FACULDADE UNYLEYA	2020	0,944521
572	UNIVERSIDADE FEDERAL FLUMINENSE	2020	0,893232
322	UNIVERSIDADE PAULISTA	2020	0,86936
663	UNIVERSIDADE SALGADO DE OLIVEIRA	2020	0,556352
1866	FACULDADE DE CIÊNCIAS EDUCACIONAIS E EMPRESARIAIS DE NATAL	2020	0,543565
14715	FACULDADE PIAGET	2020	0,530904
1084	CENTRO UNIVERSITÁRIO CNEC DE BENTO GONÇALVES	2020	0,458723
1775	CENTRO UNIVERSITÁRIO FAVIP WYDEN	2020	0,44925
1472	CENTRO UNIVERSITÁRIO LEONARDO DA VINCI	2020	0,447177
3769	FACULDADE MADRE TEREZA	2020	0,389397
1853	CENTRO UNIVERSITÁRIO UNIBTA	2020	0,317936
298	UNIVERSIDADE PITÁGORAS UNOPAR	2020	0,242817
3756	FACULDADE ALFAUNIPAC DE ALMENARA	2021	0,947525
671	UNIVERSIDADE ANHANGUERA	2021	0,884972
298	UNIVERSIDADE PITÁGORAS UNOPAR ANHANGUERA	2021	0,779288
1477	FACULDADE JK MICHELANGELO	2021	0,518903
663	UNIVERSIDADE SALGADO DE OLIVEIRA	2021	0,175605
1472	CENTRO UNIVERSITÁRIO LEONARDO DA VINCI	2021	0,160782
1853	CENTRO UNIVERSITÁRIO UNIBTA	2021	0,048288
1643	FACULDADE ANÍSIO TEIXEIRA DE FEIRA DE SANTANA	2021	0,037395
12601	FACULDADE BRASILEIRA DE ESTUDOS AVANÇADOS	2021	0,000167
4141	FACULDADE ANHANGUERA DE TAUBATÉ	2022	0,975458
370	FACULDADE PAULISTA DE ARTES	2022	0,933145
330	UNIVERSIDADE IGUAÇU	2022	0,915891
1853	CENTRO UNIVERSITÁRIO UNIBTA	2022	0,882548
569	UNIVERSIDADE FEDERAL DO PARÁ	2022	0,860299
4995	FACULDADE EDUCAMAI	2022	0,842959
527	PONTIFÍCIA UNIVERSIDADE CATÓLICA DE GOIÁS	2022	0,822489
5	UNIVERSIDADE FEDERAL DO PIAUÍ	2022	0,812266
1196	UNIVERSIDADE CESUMAR	2022	0,802932
1476	FACULDADE DE COMUNICAÇÃO E DESIGN ""OSWALDO CRUZ""	2022	0,79329
2240	FACULDADE DE FORTALEZA	2022	0,788583
3205	CENTRO UNIVERSITÁRIO DE VIÇOSA	2022	0,787536
231	FACULDADE DE SÃO BERNARDO DO CAMPO - FASB	2022	0,778032
1468	FACULDADE ESAMC UBERLÂNDIA	2022	0,776786
298	UNIVERSIDADE PITÁGORAS UNOPAR ANHANGUERA	2022	0,771951
1049	INSTITUTO DE ENSINO SUPERIOR FUCAPI	2022	0,675875

163	UNIVERSIDADE ESTÁCIO DE SÁ	2022	0,567695
18745	FACULDADE DE CIÊNCIAS APLICADAS PIAUIENSE	2022	0,529395
1712	FACULDADE DE CIÊNCIAS E TECNOLOGIA DE TERESINA	2022	0,389289
1472	CENTRO UNIVERSITÁRIO LEONARDO DA VINCI	2022	0,365358
1330	FACULDADE COTEMIG	2022	0,356499
3756	FACULDADE ALFAUNIPAC DE ALMENARA	2022	0,306779
1647	FACULDADE DE ENSINO E CULTURA DO CEARÁ	2022	0,07086
1712	FACULDADE DE CIÊNCIAS E TECNOLOGIA DE TERESINA	2023	1
1301	INSTITUTO MATONENSE MUNICIPAL DE ENSINO SUPERIOR	2023	0,998846
1785	FACULDADE DE COLIDER	2023	0,983871
14715	CENTRO UNIVERSITÁRIO PIAGET	2023	0,962516
3205	CENTRO UNIVERSITÁRIO DE VIÇOSA	2023	0,929172
3516	FACULDADE DE CIÊNCIAS DA SAÚDE ARCHIMEDES THEODORO	2023	0,92397
12268	FACULDADE DE TECNOLOGIA DE CRUZEIRO DO OESTE	2023	0,907006
1476	FACULDADE DE COMUNICAÇÃO E DESIGN ""OSWALDO CRUZ""	2023	0,900577
2623	FACULDADE DO TAPAJÓS	2023	0,867828
1196	UNIVERSIDADE CESUMAR	2023	0,825567
21583	FACULDADE DE EDUCAÇÃO SUPERIOR DE PERNAMBUCO	2023	0,812015
231	FACULDADE DE SÃO BERNARDO DO CAMPO - FASB	2023	0,797516
1898	FACULDADE DO LITORAL SUL PAULISTA	2023	0,764595
330	UNIVERSIDADE IGUAÇU	2023	0,762699
5	UNIVERSIDADE FEDERAL DO PIAUÍ	2023	0,761651
1468	FACULDADE ESAMC UBERLÂNDIA	2023	0,735183
1330	FACULDADE COTEMIG	2023	0,663661
3756	FACULDADE ALFAUNIPAC DE ALMENARA	2023	0,659639
1910	FACULDADE UNINASSAU LAURO DE FREITAS	2023	0,557692
2266	FACULDADE UNIÃO DE GOYAZES FORMOSA	2023	0,5
1049	INSTITUTO DE ENSINO SUPERIOR FUCAPI	2023	0,49534
278	FACULDADES INTEGRADAS SIMONSEN	2023	0,454495
298	UNIVERSIDADE PITÁGORAS UNOPAR ANHANGUERA	2023	0,433224
527	PONTIFÍCIA UNIVERSIDADE CATÓLICA DE GOIÁS	2023	0,417665
1472	CENTRO UNIVERSITÁRIO LEONARDO DA VINCI	2023	0,388416
13300	FACULDADE ADJETIVO CETEP	2023	0,056699
14412	FACULDADE DE TECNOLOGIA SENAI PERNAMBUCO	2023	0,035714

Fonte: Elaboração própria a partir dos microdados do Inep, 2019–2023 (Brasil, [s.d.]).

A Tabela 4 apresenta, em ordem decrescente, as IES identificadas como supereficientes na Etapa 1 ao longo do período de 2019 a 2023. No ano de 2019, destacam-se instituições como a Faculdade de São Bernardo do Campo (FASB), a Faculdade Educacional da Lapa e a Faculdade de Campina Grande do Sul, cujos escores de supereficiência se aproximam da unidade, embora permaneçam inferiores a esse limiar. Esses resultados sugerem que tais instituições, ainda que eficientes no modelo DEA clássico, apresentam vantagens marginais em relação às demais unidades eficientes quando retiradas do conjunto de referência.

Nos anos subsequentes, observa-se a manutenção de um padrão semelhante, caracterizado pela concentração dos escores em torno de valores próximos a 1 e pela ausência de casos extremos. Em 2020, por exemplo, a Faculdade de Ciência e Educação do Caparaó, a Faculdade Unyleya e a Universidade Federal Fluminense figuram entre as unidades com maior supereficiência relativa, enquanto em 2021 e 2022 destacam-se instituições como a Faculdade AlfaUnipac de Almenara, a

Universidade Anhanguera, a Faculdade Anhanguera de Taubaté e a Faculdade Paulista de Artes. Em 2023, observa-se o maior valor da série ($SE = 1$), associado à Faculdade de Ciências e Tecnologia de Teresina, seguido por instituições com escores muito próximos da unidade, como o Instituto Matonense Municipal de Ensino Superior e a Faculdade de Colíder.

A recorrência de determinadas instituições ao longo dos diferentes anos — como a Faculdade de São Bernardo do Campo (FASB), o Centro Universitário Leonardo da Vinci, a Universidade Pitágoras Unopar Anhanguera e a Pontifícia Universidade Católica de Goiás — sugere regularidade na eficiência relativa da conversão de recursos institucionais em matrículas, ainda que sem evidências de desempenho extraordinariamente superior ao conjunto.

Em síntese, os resultados da Tabela 4 indicam que a Etapa 1 do modelo NDEA apresenta baixo poder discriminante em termos de supereficiência, refletindo uma fronteira eficiente composta por unidades com desempenhos bastante próximos. A ausência de outliers positivos relevantes reforça a interpretação de que, no processo de captação de estudantes, as diferenças entre as IES eficientes decorrem de ajustes marginais na combinação de insumos, e não de vantagens estruturais acentuadas. Tal evidência justifica o aprofundamento da análise nas etapas subsequentes do modelo, particularmente aquelas associadas à retenção, conclusão e qualidade, nas quais se espera maior heterogeneidade de desempenho relativo.

A Tabela 5 apresenta as IES identificadas como supereficientes na Etapa 2 do modelo NDEA, que avalia a transformação das matrículas em concluintes, considerando o abandono estudantil como output indesejado, no período de 2019 a 2023. Essa etapa é central para a análise da eficiência educacional, pois reflete a capacidade institucional de retenção e conclusão dos estudantes.

Tabela 5 - IES supereficientes Etapa 2 (DEA-VRS OO, 2019–2023)

IES	Nome	Ano	SE_E2
19369	FACULDADE MARTINHO LUTERO	2019	1
20643	FACULDADE UNIRB - TERESINA	2019	1
1964	FACULDADE DE TECNOLOGIA EM HOTELARIA, GASTRONOMIA E TURISMO DE SP	2019	0,998679
3495	FACULDADE PINHALZINHO	2019	0,994911
527	PONTIFÍCIA UNIVERSIDADE CATÓLICA DE GOIÁS	2019	0,972528
4010	CENTRO UNIVERSITÁRIO METODISTA	2019	0,943153
952	UNIVERSIDADE SANTA CECÍLIA	2019	0,935695
4532	FACULDADE DE TECNOLOGIA SENAI CAMPO GRANDE	2019	0,927804
3285	FACULDADE SANTO ANTONIO	2019	0,926889
2554	FACULDADE DE EDUCAÇÃO SANTA TEREZINHA	2019	0,907975
1762	FACULDADE DE TECNOLOGIA E NEGÓCIOS CARLOS DRUMMOND DE ANDRADE	2019	0,904486
2160	FACULDADE SANT ANA	2019	0,879075

3769	FACULDADE MADRE TEREZA	2019	0,845969
953	UNIVERSIDADE METROPOLITANA DE SANTOS	2019	0,817247
2726	ESCOLA SUPERIOR DE ADMINISTRAÇÃO DE EMPRESAS	2019	0,801563
316	UNIVERSIDADE NOVE DE JULHO	2019	0,795819
3525	FACULDADE EVANGÉLICA DE TECNOLOGIA, CIÊNCIAS E BIOTECNOLOGIA DA CGADB	2019	0,746832
296	CENTRO UNIVERSITÁRIO DA REGIÃO DA CAMPANHA	2019	0,740292
661	FACULDADE PIO DÉCIMO	2019	0,730065
11604	FACULDADE SANTA RITA DE CHAPECÓ	2019	0,718846
1417	FACULDADE CNEC CAMPO LARGO	2019	0,629985
3294	CENTRO UNIVERSITÁRIO FAVENI	2019	0,616307
17200	FACULDADE BRASILEIRA DE TRIBUTAÇÃO	2019	0,592314
1205	FACULDADE EDUCACIONAL DA LAPA	2019	0,581086
298	UNIVERSIDADE PITÁGORAS UNOPAR	2019	0,389752
3691	FACULDADE DE TECNOLOGIA TUPY DE SÃO BENTO DO SUL	2020	1
18646	FACULDADE EVOLUÇÃO DO VALE DO ACARAÚ	2020	1
20052	FACULDADE ABC DE GOIÂNIA	2020	1
4983	FACULDADE IBRA DE TECNOLOGIA	2020	0,975846
14882	FACULDADE HORIZONTE	2020	0,970283
5	UNIVERSIDADE FEDERAL DO PIAUÍ	2020	0,968735
316	UNIVERSIDADE NOVE DE JULHO	2020	0,942337
19251	FACULDADE PÓLIS CIVITAS	2020	0,927366
4502	FACULDADES INTEGRADAS CAMPO-GRANDENSES	2020	0,921304
17326	FACULDADE GIANNA BERETTA	2020	0,894084
1401	FACULDADE ADELMAR ROSADO	2020	0,850387
13643	FACULDADES EVANGÉLICAS INTEGRADAS CANTARES DE SALOMÃO	2020	0,843481
18516	FACULDADE MIRIENSE	2020	0,830556
1756	FACULDADE ASSOCIADA BRASIL	2020	0,82666
19816	FACULDADES IMPACTOS BRASIL	2020	0,811451
1491	CENTRO UNIVERSITÁRIO INTERNACIONAL	2020	0,794946
1205	FACULDADE EDUCACIONAL DA LAPA	2020	0,765391
952	UNIVERSIDADE SANTA CECÍLIA	2020	0,752486
4	UNIVERSIDADE FEDERAL DO AMAZONAS	2020	0,649061
21095	ACADEMIA MILITAR DAS AGULHAS NEGRAS	2020	0,622989
2147	CENTRO UNIVERSITÁRIO CEUNI - FAMETRO	2020	0,596843
953	UNIVERSIDADE METROPOLITANA DE SANTOS	2020	0,557433
298	UNIVERSIDADE PITÁGORAS UNOPAR	2020	0,516139
4714	FACULDADE DE ENSINO REGIONAL ALTERNATIVA	2020	0,405918
706	FACULDADE DE ARTES DULCINA DE MORAES	2021	1
3270	FACULDADE SÃO BENTO DA BAHIA	2021	1
3782	FACULDADE DE TECNOLOGIA CACHOEIRO DE ITAPEMIRIM	2021	1
13775	FACULDADE DE TECNOLOGIA IBRATE	2021	1
14069	FACULDADE DE TECNOLOGIA PORTO SUL	2021	1
14303	FACULDADE PADRAO DE SENADOR CANEDO	2021	1
18646	FACULDADE EVOLUÇÃO DO VALE DO ACARAÚ	2021	1
18719	FACULDADE DE EDUCAÇÃO DE SANTA CATARINA	2021	1
20052	FACULDADE ABC DE GOIÂNIA	2021	1
1804	FACULDADE DE MAUÁ - FAMA	2021	0,983241
21444	FACULDADE ÁGAPE DE SÃO FÉLIX	2021	0,95
19816	FACULDADES IMPACTOS BRASIL	2021	0,946027
322	UNIVERSIDADE PAULISTA	2021	0,907481
298	UNIVERSIDADE PITÁGORAS UNOPAR ANHANGUERA	2021	0,877005
316	UNIVERSIDADE NOVE DE JULHO	2021	0,870624
17715	FACULDADE DE EDUCAÇÃO SUPERIOR DE PERNAMBUCO	2021	0,82427
18745	FACULDADE DE CIÊNCIAS APLICADAS PIAUIENSE	2021	0,817427
135	CLARETIANO - CENTRO UNIVERSITÁRIO	2021	0,711882
3525	FACULDADE EVANGÉLICA DE TECNOLOGIA, CIÊNCIAS E BIOTECNOLOGIA DA CGADB	2021	0,38087
17115	FACULDADE TRÊS MARIAS	2021	0,262455
1205	CENTRO UNIVERSITÁRIO FAEL	2021	0,131091
1585	FACULDADE ENERGIA	2022	1
1653	FACULDADE DE CIÊNCIA E EDUCAÇÃO DO CAPARAÓ	2022	1
2726	FACULDADE BILAC DE SÃO JOSÉ DOS CAMPOS	2022	1
3270	FACULDADE SÃO BENTO DA BAHIA	2022	1

3692	FACULDADE DE TECNOLOGIA IPEP DE CAMPINAS	2022	1
5133	FACULDADE SENAC SÃO MIGUEL DO OESTE	2022	1
13034	FACULDADE DE ENSINO SUPERIOR	2022	1
14303	FACULDADE PADRAO DE SENADOR CANEDO	2022	1
14858	FACULDADE DE TECNOLOGIA FRANCISCO MORATO	2022	1
14879	FACULDADE DOURADO	2022	1
18072	FACULDADE DE DIREITO TECH DE SÃO PAULO	2022	1
19206	FACULDADE UNIRB - FORTALEZA	2022	1
20052	FACULDADE ABC DE GOIÂNIA	2022	1
2343	CENTRO UNIVERSITÁRIO U:VERSE	2022	0,972284
4714	FACULDADE DE ENSINO REGIONAL ALTERNATIVA	2022	0,945493
452	CENTRO UNIVERSITÁRIO LUTERANO DE MANAUS	2022	0,93043
163	UNIVERSIDADE ESTÁCIO DE SÁ	2022	0,909098
322	UNIVERSIDADE PAULISTA	2022	0,902332
1541	FACULDADE FLAMINGO	2022	0,877863
2745	CENTRO UNIVERSITÁRIO DA AMAZÔNIA	2022	0,815953
3164	INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA DE MATO GROSSO	2022	0,811601
17025	FACULDADE DE EDUCAÇÃO SUPERIOR DE PARAGOMINAS	2022	0,746284
2575	FACULDADES INTEGRADAS SOARES DE OLIVEIRA	2022	0,745236
298	UNIVERSIDADE PITÁGORAS UNOPAR ANHANGUERA	2022	0,705892
135	CLARETIANO - CENTRO UNIVERSITÁRIO	2022	0,633404
316	UNIVERSIDADE NOVE DE JULHO	2022	0,631315
2147	CENTRO UNIVERSITÁRIO CEUNI - FAMETRO	2022	0,571574
671	UNIVERSIDADE ANHANGUERA	2022	0,558462
15859	FACULDADE CATÓLICA PAULISTA	2022	0,385694
216	CENTRO UNIVERSITÁRIO METODISTA IZABELA HENDRIX	2023	1
299	FACULDADE DE JANDAIA DO SUL	2023	1
336	FACULDADE MACHADO SOBRINHO	2023	1
607	FAC. DE CIÊNCIAS CONTÁBEIS E ADMINISTRATIVAS DE CACHOEIRO DO ITAPEMIRIM	2023	1
615	FAC. DE FILOSOFIA CIÊNCIAS E LETRAS PROFESSORA NAIR FORTES ABU-MERHY	2023	1
1090	FACULDADE SÃO LUÍS DE FRANÇA	2023	1
1385	FACULDADES INTEGRADAS IPEP	2023	1
1477	FACULDADE ERICH FROMM	2023	1
1562	FACULDADE DE ILHA SOLTEIRA	2023	1
1682	FACULDADE CATÓLICA SALESIANA DE MACAÉ	2023	1
1910	FACULDADE UNINASSAU LAURO DE FREITAS	2023	1
1964	FACULDADE DE TECNOLOGIA EM HOTELARIA, GASTRONOMIA E TURISMO DE SP	2023	1
2009	FACULDADE DE TAQUARITINGA	2023	1
2027	FACULDADE DE CAMPINA GRANDE	2023	1
2165	FACULDADE FOZ DO IGUAÇU	2023	1
2246	FACULDADE METROPOLITANA DE CIENCIAS E TECNOLOGIA	2023	1
2266	FACULDADE UNIÃO DE GOYAZES FORMOSA	2023	1
2879	FACULDADE DE CAIEIRAS	2023	1
2903	FACULDADE CONCÓRDIA	2023	1
3008	FAC. DO INSTITUTO NACIONAL DE PÓS-GRADUAÇÃO DE SÃO JOSÉ DOS CAMPOS	2023	1
3302	FACULDADE REGES DE JUNQUEIRÓPOLIS	2023	1
3306	FACULDADE DE TECNOLOGIA IPUC	2023	1
3448	FACULDADE IBITURUNA	2023	1
3525	FACULDADE EVANGÉLICA DE TECNOLOGIA, CIÊNCIAS E BIOTECNOLOGIA DA CGADB	2023	1
3692	FACULDADE DE TECNOLOGIA IPEP DE CAMPINAS	2023	1
3837	FACULDADE EVOLUÇÃO	2023	1
4218	FACULDADE PEDRO LEOPOLDO	2023	1
4357	FACULDADE QUALITTAS	2023	1
4443	FACULDADE MERIDIONAL DE IJUÍ	2023	1
4714	FACULDADE DE ENSINO REGIONAL ALTERNATIVA	2023	1
5000	FACULDADE ÁGORA - ADMINISTRAÇÃO, EDUCAÇÃO E CULTURA	2023	1
11586	FACULDADES NETWORK - CAMPUS SUMARÉ	2023	1
12338	FACULDADE DE TECNOLOGIA LA SALLE - ESTRELA	2023	1
13775	FACULDADE DE TECNOLOGIA IBRATE	2023	1
14069	FACULDADE DE TECNOLOGIA PORTO SUL	2023	1
14242	FACULDADE CONEXÃO	2023	1
14412	FACULDADE DE TECNOLOGIA SENAI PERNAMBUCO	2023	1

14858	FACULDADE DE TECNOLOGIA FRANCISCO MORATO	2023	1
15428	FACULDADE UNIRB - NATAL	2023	1
15526	FACULDADE UNIRB CEARÁ	2023	1
16502	FACULDADE AETOS	2023	1
17490	FACULDADE JOÃO PAULO II	2023	1
17666	FACULDADE ANHANGUERA UNIDADE PARAGOMINAS	2023	1
17688	FACULDADE FADAM	2023	1
18019	FACULDADE DO EDUCADOR	2023	1
18072	FACULDADE DE DIREITO TECH DE SÃO PAULO	2023	1
18659	FACULDADE ANHANGUERA DO CEARÁ	2023	1
19206	FACULDADE UNIRB - FORTALEZA	2023	1
19219	FACULDADE FASIPE DF	2023	1
19793	FACULDADE FADAM DE MARACANAÚ	2023	1
19853	FACULDADE DOM BOSCO PARANÁ	2023	1
20052	FACULDADE ABC DE GOIÂNIA	2023	1
20666	FACULDADE ELO	2023	1
21215	FACULDADE DE ENSINO DE PESQUISA DO COOPERATIVISMO	2023	1
21463	FACULDADE DE ENSINO PORTAL	2023	1
1294	CENTRO UNIVERSITÁRIO DAS AMÉRICAS	2023	0,996587
2186	FACULDADE SALESIANA DOM BOSCO	2023	0,978349
802	CENTRO UNIVERSITÁRIO UNICARIOCA	2023	0,977585
1805	CENTRO UNIVERSITÁRIO DOS GUARARAPES	2023	0,966181
18009	FACULDADE BSSP	2023	0,963401
135	CLARETIANO - CENTRO UNIVERSITÁRIO	2023	0,96165
322	UNIVERSIDADE PAULISTA	2023	0,933248
278	FACULDADES INTEGRADAS SIMONSEN	2023	0,920177
569	UNIVERSIDADE FEDERAL DO PARÁ	2023	0,868006
952	UNIVERSIDADE SANTA CECÍLIA	2023	0,862151
712	CENTRO UNIVERSITÁRIO IBMR	2023	0,860274
2311	FACULDADE KENNEDY DE MINAS GERAIS - FKMG	2023	0,82189
2832	FACULDADE SUCESSO	2023	0,63843
298	UNIVERSIDADE PITÁGORAS UNOPAR ANHANGUERA	2023	0,456768
4751	CENTRO UNIVERSITÁRIO UNIFATECIE	2023	0,439181
316	UNIVERSIDADE NOVE DE JULHO	2023	0,364668
898	CENTRO UNIVERSITÁRIO UNISAN	2023	0,158654

Fonte: Elaboração própria a partir dos microdados do Inep, 2019–2023 (Brasil, [s.d.]).

Os resultados indicam heterogeneidade temporal no comportamento da supereficiência. Em 2019 e 2020, observa-se um conjunto relativamente restrito de instituições com supereficiência elevada, incluindo casos com $SE = 1$, seguidos por valores próximos da unidade, o que sugere boa capacidade discriminante da fronteira nesse período. Destacam-se, nesses anos, instituições como a Faculdade Martinho Lutero, a Faculdade UNIRB – Teresina e a Faculdade de Tecnologia Tupy de São Bento do Sul.

A partir de 2021, verifica-se um aumento expressivo no número de IES com supereficiência unitária, padrão que se intensifica em 2022 e 2023. Esse comportamento indica uma fronteira eficiente mais ampla e menos sensível à exclusão individual das DMUs, reduzindo o poder discriminante do modelo nesse estágio. Assim, valores de $SE = 1$ devem ser interpretados menos como desempenho excepcional e mais como evidência de homogeneidade tecnológica no processo de retenção e conclusão.

De forma geral, os resultados da Etapa 2 sugerem que, embora algumas instituições apresentem desempenho recorrente ao longo do período, a supereficiência assume predominantemente um papel diagnóstico, contribuindo para a identificação de padrões estruturais da fronteira, e não para o ranqueamento estrito das IES. Essa constatação reforça a importância da análise da Etapa 3, voltada aos resultados institucionais e à dimensão externa da eficiência.

A Tabela 6 apresenta as IES supereficientes na Etapa 3 do modelo NDEA entre 2019 e 2023, etapa que avalia a conversão dos concluintes em resultados institucionais finais, incorporando indicadores de qualidade acadêmica. Observa-se que os escores de supereficiência concentram-se próximos da unidade, com menor dispersão em relação às etapas anteriores, indicando maior homogeneidade tecnológica e estabilidade na fronteira associada aos resultados.

Tabela 6 - IES supereficientes Etapa 3 (DEA-VRS OO, 2019–2023)

IES	Nome	Ano	SE_E3
2591	EPGE ESCOLA BRASILEIRA DE ECONOMIA E FINANÇAS	2019	0,986316
3614	ESCOLA DE CIÊNCIAS SOCIAIS	2019	0,978791
21952	FGV / ESCOLA DE RELAÇÕES INTERNACIONAIS	2019	0,865754
3182	FACULDADE UNIRB - ARAPIRACA	2019	0,861517
2111	CENTRO UNIVERSITÁRIO INTA	2019	0,524619
1472	CENTRO UNIVERSITÁRIO LEONARDO DA VINCI	2019	0,105226
17403	FACULDADE ARI DE SÁ	2019	0,027534
18979	FACULDADE MULTIVIX VILA VELHA	2020	0,991867
298	UNIVERSIDADE PITÁGORAS UNOPAR	2020	0,967167
3614	ESCOLA DE CIÊNCIAS SOCIAIS	2020	0,946116
227	UNIVERSIDADE CATÓLICA DE SANTOS	2020	0,894317
19327	FACULDADE CESAR	2020	0,88398
21952	FGV / ESCOLA DE RELAÇÕES INTERNACIONAIS	2020	0,879047
1510	CENTRO UNIVERSITÁRIO ESTÁCIO DE SANTA CATARINA - ESTÁCIO SANTA CATARINA	2020	0,699428
2189	FACULDADE DO MARANHÃO	2020	0,570947
19607	FACULDADE SANTÍSSIMA TRINDADE	2020	0,534513
1472	CENTRO UNIVERSITÁRIO LEONARDO DA VINCI	2020	0,497622
3614	ESCOLA DE CIÊNCIAS SOCIAIS FGV CPDOC	2021	0,931388
21952	FGV / ESCOLA DE RELAÇÕES INTERNACIONAIS	2021	0,611021
589	UNIVERSIDADE FEDERAL RURAL DO SEMI-ÁRIDO	2021	0,120602
18684	FACULDADE SANTA TERESA	2021	0,101201
1472	CENTRO UNIVERSITÁRIO LEONARDO DA VINCI	2021	0,01012
1851	ESCOLA BRASILEIRA DE ADMINISTRAÇÃO PÚBLICA E DE EMPRESAS	2022	0,999359
22197	FACULDADE ATENAS PASSOS	2022	0,995641
2591	EPGE ESCOLA BRASILEIRA DE ECONOMIA E FINANÇAS	2022	0,992311
22194	FACULDADE ATENAS SETE LAGOAS	2022	0,966741
21554	FACULDADE ANHANGUERA DE ITAPEVA	2022	0,932665
3614	ESCOLA DE CIÊNCIAS SOCIAIS FGV CPDOC	2022	0,891293
926	CENTRO UNIVERSITÁRIO ANHANGUERA PITÁGORAS UNOPAR DE CAMPO GRANDE	2022	0,877473
3756	FACULDADE ALFAUNIPAC DE ALMENARA	2022	0,799261
22917	FACULDADE DE CIÊNCIAS MÉDICAS DE MARICÁ	2022	0,747182
1472	CENTRO UNIVERSITÁRIO LEONARDO DA VINCI	2022	0,018434
1851	ESCOLA BRASILEIRA DE ADMINISTRAÇÃO PÚBLICA E DE EMPRESAS	2023	0,991511
2688	FACULDADE INESP - INSTITUTO NACIONAL DE ENSINO E PESQUISA	2023	0,988354
1660	CENTRO UNIVERSITÁRIO ESTÁCIO DE BRASÍLIA - ESTÁCIO BRASÍLIA	2023	0,959567
3614	ESCOLA DE CIÊNCIAS SOCIAIS FGV CPDOC	2023	0,783682

19325	FACULDADE DO CENTRO OESTE PAULISTA	2023	0,764053
3756	FACULDADE ALFAUNIPAC DE ALMENARA	2023	0,608975
23318	FACULDADE BRASILEIRA DE CACHOEIRO	2023	0,326648
1472	CENTRO UNIVERSITÁRIO LEONARDO DA VINCI	2023	0,02357

Fonte: Elaboração própria a partir dos microdados do Inep, 2019–2023 (Brasil, [s.d.]).

Destacam-se instituições com desempenho recorrente, como a EPGE Escola Brasileira de Economia e Finanças e a Escola de Ciências Sociais (FGV/CPDOC), sugerindo consistência na entrega de resultados acadêmicos. Diferentemente das etapas anteriores, não se verificam valores extremos, o que reforça a interpretação de que os indicadores finais são menos sensíveis a variações abruptas. Em sentido oposto, escores reduzidos observados para algumas instituições indicam maior distanciamento da fronteira eficiente, ainda que classificadas como eficientes no DEA-VRS tradicional.

A superineficiência constitui um recurso metodológico complementar para identificar instituições que, embora classificadas como eficientes pelo DEA-VRS, apresentam desempenhos relativamente inferiores quando avaliadas em comparação com a fronteira invertida. Esse procedimento estima supereficiência em combinação com a fronteira invertida e permite destacar outliers negativos, ou seja, unidades que demandariam acréscimos proporcionais em seus outputs para alcançarem níveis de desempenho compatíveis com as demais instituições eficientes (Pastor *et al.*, 1999). Avalia o grau de proximidade da DMU em relação às práticas menos eficientes identificadas.

A partir dos resultados de supereficiência obtidos no modelo VRS, tornou-se necessário proceder ao teste estatístico de comparação entre modelos, de modo a verificar a adequação entre as hipóteses de retornos constantes e variáveis de escala. Para esse fim, adotou-se o teste de Kolmogorov-Smirnov (KS), aplicado às distribuições de eficiência, permitindo avaliar a escolha do modelo mais apropriado ao conjunto de dados analisado.

4.3 TESTE DE MODELO (CRS OU VRS)

O procedimento adotado nesta etapa consistiu na realização do teste de Kolmogorov-Smirnov (KS), com o objetivo de comparar as distribuições dos escores de eficiência obtidos pelos modelos CRS (Constant Returns to Scale) e VRS (Variable Returns to Scale) em cada ano do período de 2019 a 2023. A hipótese nula (H_0) estabelece que as duas distribuições são estatisticamente iguais, o que implicaria a

adoção do modelo CRS. A hipótese alternativa (H_1), por sua vez, indica que as distribuições diferem, justificando a utilização do modelo VRS.

Os resultados do teste estão apresentados na Tabela 7, que mostra a razão entre as somas dos escores de eficiência, a estatística KS, os valores-p correspondentes e o modelo adequado identificado em cada ano.

Tabela 7 - Resultados do teste KS para os anos de 2019 a 2023

Ano	Ratio	KS_Estatistica	KS_p_valor	Modelo_Adequado
2019	4,51980965	0,66312997	1,6841E-21	VRS
2020	1,38242592	0,58938992	1,3994E-18	VRS
2021	1,52880772	0,64809903	1,4755E-47	VRS
2022	1,57677139	0,65658709	2,6584E-33	VRS
2023	1,91519543	0,68505747	2,9174E-31	VRS

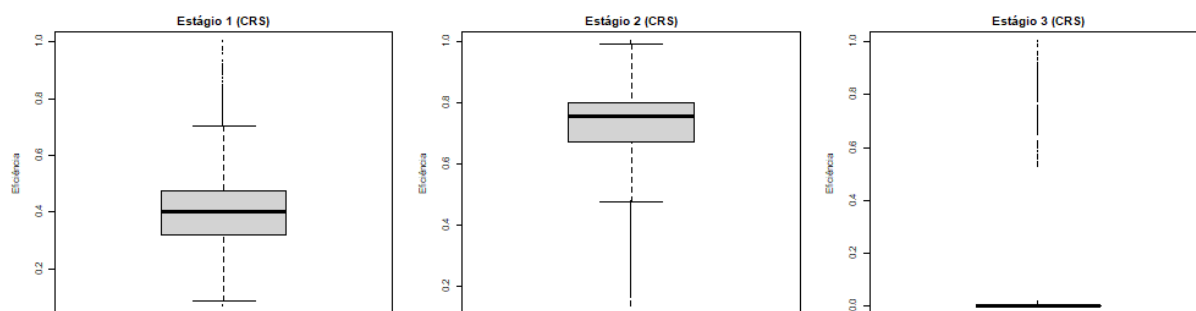
Fonte: Elaboração própria a partir dos resultados do teste KS.

A análise dos resultados revela que, em todos os anos, os valores-p foram inferiores a 0,05, indicando diferenças estatisticamente significativas entre as distribuições de eficiências obtidas nos modelos CRS e VRS. Além disso, os valores da estatística KS variaram entre 0,58 e 0,68, confirmando a magnitude da discrepância entre os modelos. Esse resultado é consistente com o valor da razão dos escores de eficiência, que apresentou índices superiores a 1 em todos os anos, sugerindo que o CRS subestima a eficiência relativa quando comparado ao VRS.

Além do teste estatístico de Kolmogorov-Smirnov, a análise gráfica por meio dos boxplots das distribuições de eficiência reforça as diferenças entre os modelos CRS e VRS. Nos gráficos referentes ao modelo CRS, observa-se maior dispersão nos escores de eficiência, com destaque para a terceira etapa, em que há forte concentração de valores próximos de zero e presença de outliers, o que indica subestimação da performance das instituições quando se supõe retornos constantes de escala. Já nos boxplots do modelo VRS, a distribuição apresenta-se mais equilibrada, com medianas mais elevadas e menor assimetria, sobretudo na segunda e terceira etapas, sugerindo que o modelo com retornos variáveis de escala é mais adequado para capturar a heterogeneidade das IES. Essa diferença visual complementa os achados do teste KS, evidenciando que a imposição de retornos constantes tende a distorcer a avaliação da eficiência relativa, enquanto o modelo VRS confere maior robustez e aderência à realidade observada no setor.

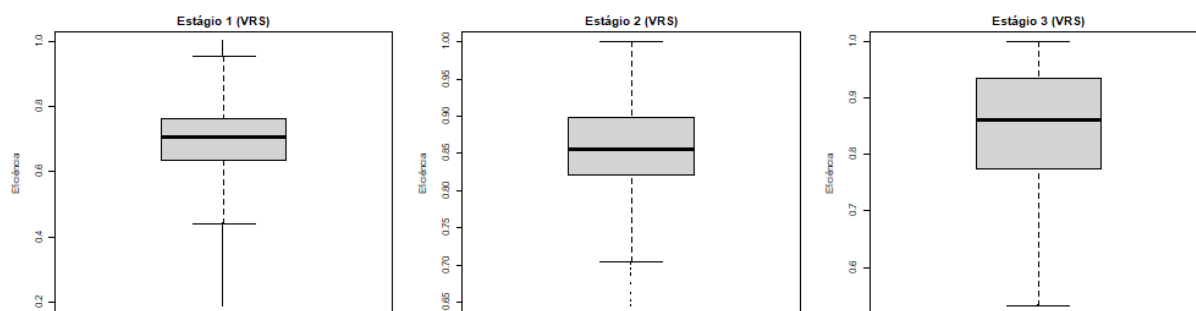
Conclui-se, portanto, que o modelo VRS representa de forma mais consistente as tecnologias das IES avaliadas, assegurando maior aderência à realidade marcada por heterogeneidade estrutural e de escala.

Gráfico 3 - Boxplots das Eficiências por Etapas (CRS)



Fonte: Elaboração própria a partir dos resultados do teste KS.

Gráfico 4 - Boxplots das Eficiências por Etapas (VRS)



Fonte: Elaboração própria a partir dos resultados do teste KS.

4.4 ANÁLISE DE GRUPOS (IES PÚBLICAS x PRIVADAS)

Com a definição do modelo de referência, torna-se possível avançar para a análise comparativa entre grupos de instituições, de modo a investigar se as diferenças entre IES públicas e privadas impactam seus níveis de eficiência.

Para verificar se a natureza institucional das IES influencia os níveis de eficiência, foi aplicado o teste de Kruskal-Wallis (KW), um teste estatístico não paramétrico que compara distribuições de grupos independentes sem pressupor normalidade. A comparação envolveu IES públicas e privadas em cada ano da série 2019–2023, tomando em consideração as hipóteses:

H_0 (hipótese nula): as populações associadas aos k grupos têm a mesma distribuição;

H_1 (alternativa): pelo menos um grupo difere dos demais;

Sob H_0 , segue aproximadamente uma distribuição qui-quadrado com $k-1$ graus de liberdade. Se $p \leq \alpha$: rejeita-se $H_0 \rightarrow$ um grupo difere do outro.

Os resultados, apresentados na Tabela 8, indicam valores da estatística qui-quadrado elevados em todos os anos, com p-valores inferiores a 0,05. Em 2019, a estatística foi de 91,63 (indicando um $p < 0,001$); em 2020, de 83,17 (indicando um $p < 0,001$); em 2021, de 104,48 (indicando um $p < 0,001$); em 2022, de 98,45 (indicando um $p < 0,001$); e em 2023, de 79,80 (indicando um $p < 0,001$). No consolidado, o resultado foi ainda mais expressivo $\text{Chi_Sq} = 453,07$; indicando um $p < 0,001$), confirmando a significância global do teste.

Tabela 8 - Resultados do teste KW para os anos de 2019 a 2023

Ano	Chi_Sq	DF	p_valor	Significativo
2019	91,6369969	1	1,0413E-21	Sim
2020	83,1749233	1	7,5104E-20	Sim
2021	104,484304	1	1,5844E-24	Sim
2022	98,4511957	1	3,3313E-23	Sim
2023	79,8037627	1	4,135E-19	Sim
Todos	453,072311	1	1,547E-100	Sim

Fonte: Elaboração própria a partir dos resultados do teste KW.

Esses resultados demonstram que há diferenças estatisticamente significativas entre IES públicas e privadas em todos os anos avaliados, bem como no consolidado. Tal evidência reforça que a estrutura institucional exerce influência direta sobre os níveis de eficiência observados, o que pode estar relacionado a diferenças de porte, regime de financiamento, políticas de expansão e estratégias de gestão. Dessa forma, a análise de grupos confirma que a heterogeneidade entre redes não apenas se reflete na escala de operação, mas também impacta de forma consistente o desempenho relativo das instituições ao longo do período analisado.

Em síntese, o teste de Kruskal-Wallis evidenciou diferenças estatisticamente significativas entre IES públicas e privadas, confirmando a influência da natureza institucional nos níveis de eficiência e reforçando a adoção do modelo VRS. Concluída essa etapa, passa-se ao NDEA-clássico e ao bootstrap orientado a outputs, a fim de avaliar a consistência estatística dos escores e reduzir potenciais vieses.

4.5 NDEA-CLÁSSICO E BOOTSTRAP ORIENTADO AOS PRODUTOS

Antes da apresentação e discussão dos resultados, é importante destacar que as análises do DEA-clássico e do DEA-bootstrap orientado a outputs, na primeira etapa, bem como a avaliação do impacto de fatores exógenos em etapas posteriores, foram realizadas a partir do conjunto consolidado das informações prestadas ao MEC pelas Instituições de Ensino Superior no período de 2019 a 2023.

Para os cálculos do NDEA nas três etapas, adotou-se o modelo VRS com orientação a outputs, conforme validado no teste de escolha de fronteira anteriormente realizado. A decisão encontra respaldo em trabalhos como os de Brzezicki (2020), Civera *et al.* (2020) e Barra e Zotti (2016), que aplicaram o DEA-VRS associado ao procedimento bootstrap para mensuração da eficiência em departamentos ou instituições de ensino superior. No caso do estudo de Barra e Zotti (2016), destaca-se ainda o uso do índice de Malmquist para avaliar a dinâmica da produtividade, o que reforça a pertinência metodológica deste trabalho, dado o paralelismo com as técnicas aqui adotadas.

Adicionalmente, ao adotar modelos VRS, introduz-se nos cálculos de eficiência um fator escala inerente ao porte das IES. Ou seja, espera-se que instituições maiores apresentem margens mais favoráveis de eficiência, exatamente pela sua capacidade de diluir custos fixos e aproveitar economias de escala. Essa hipótese encontra respaldo em estudos anteriores, como Lee (2011), Martínez-Campillo e Fernández-Santos (2020) e Agasisti *et al.* (2021), que demonstraram empiricamente efeitos positivos do tamanho institucional sobre os escores de eficiência em contextos educacionais ou acadêmicos.

Com a definição do modelo, a etapa seguinte consistiu no cálculo da eficiência em três etapas distintas por meio da aplicação do Network DEA, metodologia também adotada em estudos anteriores, como na avaliação de programas de pós-graduação em Economia (Falchetto *et al.*, 2018) e em Engenharias (Gomes Júnior *et al.*, 2014). Em cada etapa é estimada uma medida de eficiência própria, e, a partir dessas, procede-se ao cálculo da eficiência global, obtida pela multiplicação dos índices de eficiência de cada etapa do processo produtivo acadêmico.

Na etapa 1 (Eficiência de Captação) são avaliadas as condições iniciais de funcionamento das instituições. Nessa fase, foram considerados como insumos o número de docentes e de técnicos administrativos, bem como os serviços ofertados e

o acervo disponibilizado. Como produto, foi utilizado o total de alunos matriculados em cada IES. A ideia é que instituições capazes de atender maior número de estudantes com estrutura proporcionalmente menor tendem a apresentar maior eficiência operacional na etapa inicial do processo produtivo.

A etapa 2 (Eficiência Interna) busca mensurar a eficiência pedagógica na transformação das matrículas em resultados acadêmicos. Para tanto, a entrada considerada foi o total de alunos matriculados, e como saídas foram incluídos os concluintes e, adicionalmente, o abandono como produto indesejado. A inclusão dessa variável negativa é fundamental, uma vez que a evasão representa desperdício de recursos e impacto direto sobre a eficiência. O objetivo dessa etapa é, portanto, maximizar a titulação dos estudantes e minimizar as perdas associadas aos desligamentos.

Por fim, a etapa 3 (Eficiência Externa) avalia o desempenho final das IES, considerando como entrada o número de alunos concluintes e como saídas indicadores de qualidade institucional, representados pelo Índice Geral de Cursos (IGC) e pelos indicadores de satisfação dos discentes. Essa etapa captura a capacidade das instituições em traduzir os esforços anteriores em resultados reconhecidos socialmente e em valor agregado à formação acadêmica.

Dessa forma, os três subprocessos do NDEA permitem examinar desde a captação de recursos humanos e estruturais até a geração de resultados finais de qualidade. A análise foi conduzida tanto pelo método clássico (determinístico) quanto pelo bootstrap orientado a outputs, o que possibilitará, em momento posterior, a comparação entre os resultados e a avaliação da robustez das estimativas de eficiência.

A análise das trinta Instituições de Ensino Superior mais bem classificadas no NDEA clássico, apresentada na Tabela 9, evidencia um predomínio expressivo das instituições privadas no topo do ranking de eficiência. Observa-se que quatro instituições privadas – Universidade Nove de Julho, Universidade Paulista, Centro Universitário de Maringá (Unicesumar) e Faculdade ABC de Goiânia – atingiram eficiência máxima em todas as etapas do processo produtivo acadêmico (E1, E2 e E3), alcançando eficiência global igual a 1. Esse resultado ilustra a capacidade dessas instituições em equilibrar de forma consistente a alocação de recursos, a retenção de estudantes e a geração de resultados acadêmicos finais, mantendo desempenho robusto em todo o ciclo.

Tabela 9 - Ranking de Eficiências NDEA Clássico (2023)

Nº	DMU	Nome	Rede	E1	E2	E3	EG
1	316	Universidade Nove De Julho	Privada	1	1	1	1
2	322	Universidade Paulista	Privada	1	1	1	1
3	1196	Centro Universitário De Maringá	Privada	1	1	1	1
4	20052	Faculdade ABC De Goiânia	Privada	1	1	1	1
5	466	Universidade Anhembi Morumbi	Privada	0,97664	0,986054	1	0,963019
6	1491	Centro Univ. Internacional	Privada	1	1	0,935754	0,935754
7	1472	Centro Univ. Leonardo Da Vinci	Privada	1	1	0,935754	0,935754
8	1910	Fac. Uninassau Lauro De Freitas	Privada	1	1	0,893851	0,893851
9	3840	Centro Universitário Facvest	Privada	0,962061	0,920035	1	0,88513
10	203	Universidade São Judas Tadeu	Privada	0,913896	0,944191	1	0,862892
11	1977	Centro De Estudos Superiores Aprendiz	Privada	1	1	0,861588	0,861588
12	298	Universidade Pitágoras Unopar	Privada	1	1	0,861588	0,861588
13	569	Universidade Federal Do Pará	Pública	0,892148	0,963586	1	0,859661
14	4751	Fac. Tecnol E Ciencias Do Norte Do Paraná	Privada	1	0,991327	0,861588	0,854115
15	338	Pontifícia Univ. Católica De Minas Gerais	Privada	0,905249	0,943024	1	0,853671
16	221	Universidade Cruzeiro Do Sul	Privada	1	0,986415	0,861588	0,849883
17	2	Universidade De Brasília	Pública	0,901543	0,941129	1	0,848468
18	712	Centro Universitário Ibmr	Privada	0,906068	1	0,935754	0,847857
19	586	Universidade Federal Do Rio De Janeiro	Pública	0,915898	0,9182	1	0,840977
20	12268	Fac. De Tecnologia De Cruzeiro Do Oeste	Privada	0,972587	1	0,861588	0,837969
21	385	Universidade Salvador	Privada	0,864914	0,96833	1	0,837522
22	448	Centro Universitário Ritter Dos Reis	Privada	0,867203	0,964686	1	0,836578
23	1805	Centro Universitário Dos Guararapes	Privada	0,852612	0,980469	1	0,835959
24	718	Universidade Potiguar	Privada	0,875555	0,954692	1	0,835886
25	1205	Faculdade Educacional Da Lapa	Privada	0,963046	0,960466	0,893851	0,826788
26	18637	Faculdade Imesmercosur	Privada	0,980249	0,899497	0,935754	0,825083
27	575	Universidade Federal De Minas Gerais	Pública	0,884916	0,929242	1	0,822301
28	417	Universidade Cidade De São Paulo	Privada	0,98845	0,965498	0,861588	0,822253
29	1396	Centro Universitário Dinâmica Das Cataratas	Privada	0,949457	0,864036	1	0,820365
30	374	Centro Univ. Das Fac. Metropolitanas Unidas	Privada	0,9338	0,938599	0,935754	0,820154

Fonte: Elaboração própria no R-Studio.

Ainda entre as privadas, várias IES mantiveram escores próximos da eficiência máxima, como a Universidade Anhembi Morumbi, com desempenho pleno na etapa externa (E3), mas pequenas perdas nas etapas iniciais; e os Centros Universitários Internacional, Leonardo da Vinci e Uninassau Lauro de Freitas, que apresentaram resultados elevados nas fases de captação e interna, mas eficiência reduzida na etapa final. Essa redução reflete uma dificuldade em converter integralmente os concluintes em indicadores de qualidade e satisfação, sugerindo limites no desempenho final, mesmo quando a base operacional e pedagógica é sólida.

No que se refere às instituições públicas, verifica-se a presença de quatro universidades federais de grande porte entre as 30 mais eficientes: Universidade Federal do Pará (UFPA), Universidade de Brasília (UnB), Universidade Federal do Rio de Janeiro (UFRJ) e Universidade Federal de Minas Gerais (UFMG). Todas elas apresentaram valores elevados de eficiência global, acima de 0,82, com destaque para a UFPA, que alcançou desempenho máximo na etapa externa. Isso demonstra que, apesar de não estarem na liderança do ranking, as universidades públicas

federais apresentam consistência nos resultados finais, sinalizando a relevância de seu papel na formação acadêmica e na geração de qualidade educacional.

Outro aspecto a ser destacado é que, mesmo entre as IES mais bem posicionadas, poucas conseguiram manter eficiência total nas três etapas simultaneamente. A maioria das instituições apresentou eficiência máxima em uma ou duas fases, mas não no ciclo completo, o que reduziu a eficiência global. Esse comportamento indica que alcançar eficiência plena em captação, retenção e resultados externos é um desafio significativo, sendo mais comum que as instituições se destaquem em etapas específicas.

Em síntese, os resultados do NDEA clássico confirmam uma clara predominância das instituições privadas entre as mais eficientes, mas também evidenciam que universidades públicas federais consolidadas ocupam posição relevante no ranking, reforçando sua importância acadêmica e científica. A diversidade de perfis institucionais presentes entre as 30 mais bem classificadas demonstra que tanto, instituições de grande porte quanto centros universitários privados, conseguem atingir altos níveis de eficiência, ainda que por caminhos distintos. Esses achados fornecem uma primeira visão do desempenho das IES brasileiras no último ano analisado e servirão de base para a etapa seguinte, em que os resultados serão reavaliados sob o procedimento de bootstrap orientado a outputs, a fim de verificar sua robustez.

4.5.1 Cálculo da Eficiência a partir da utilização do bootstrap

A aplicação do NDEA-bootstrap proporcionou uma avaliação mais conservadora e robusta das eficiências, corrigindo possíveis vieses presentes nos resultados determinísticos. Como esperado, os escores de eficiência global apresentaram redução em comparação com o NDEA clássico, sobretudo nos casos em que anteriormente havia prevalência de valores iguais a 1. Esse ajuste reflete a maior precisão estatística proporcionada pelo procedimento, que reposiciona as instituições mais próximas da fronteira eficiente sem, contudo, superestimar sua performance.

Na Tabela 10, observa-se novamente a predominância das instituições privadas entre as primeiras posições do ranking. A Universidade Paulista, o Centro Universitário de Maringá (Unicesumar) e a Universidade Nove de Julho mantêm-se no topo, ainda que com escores ligeiramente inferiores aos observados no NDEA

clássico. Outras privadas de grande porte, como a Universidade Anhembi Morumbi e os Centros Universitários Internacional e Leonardo da Vinci, também se destacam entre as dez primeiras colocadas, confirmando o padrão de elevada eficiência operacional e acadêmica nessas instituições.

Tabela 10 - Ranking de Eficiências NDEA Bootstrap (2023)

Nº	DMU	Nome	Rede	E1	E2	E3	EG
1	322	Universidade Paulista	Privada	1	0,999946	1	0,999946
2	1196	Centro Universitário De Maringá - Unicesumar	Privada	0,999883	0,999946	1	0,999829
3	316	Universidade Nove De Julho	Privada	0,999866	0,986212	1	0,986081
4	466	Universidade Anhembi Morumbi	Privada	0,976197	0,985812	1	0,962346
5	1491	Centro Universitário Internacional	Privada	0,999191	0,999946	0,935754	0,934947
6	1472	Centro Universitário Leonardo Da Vinci	Privada	0,98969	0,999946	0,935754	0,926056
7	203	Universidade São Judas Tadeu	Privada	0,909282	0,942114	1	0,856647
8	221	Universidade Cruzeiro Do Sul	Privada	0,99926	0,986362	0,861588	0,849208
9	338	Pontifícia Universidade Católica De Minas Gerais	Privada	0,905245	0,937236	1	0,848428
10	298	Universidade Pitágoras Unopar	Privada	0,99626	0,999946	0,85078	0,847552
11	1910	Faculdade Uninassau Lauro De Freitas	Privada	0,947682	0,999945	0,884382	0,838067
12	712	Centro Universitário Ibm	Privada	0,891853	0,998981	0,935754	0,833705
13	385	Universidade Salvador	Privada	0,8602	0,966993	1	0,831807
14	718	Universidade Potiguar	Privada	0,871293	0,953542	1	0,830815
15	2	Universidade De Brasília	Pública	0,901543	0,91894	1	0,828464
16	448	Centro Universitário Ritter Dos Reis	Privada	0,855754	0,963122	1	0,824196
17	569	Universidade Federal Do Pará	Pública	0,892139	0,922155	1	0,822691
18	4751	Fac. Tecnologia E Ciencias Do Norte Do Paraná	Privada	0,981449	0,970225	0,861588	0,820427
19	417	Universidade Cidade De São Paulo	Privada	0,985063	0,965205	0,861588	0,819187
20	374	Centro Univ. Das Fac. Metropolitanas Unidas	Privada	0,933109	0,938073	0,935754	0,819089
21	1805	Centro Universitário Dos Guararapes	Privada	0,837632	0,977616	1	0,818882
22	586	Universidade Federal Do Rio De Janeiro	Pública	0,915898	0,890731	1	0,815818
23	3840	Centro Universitário Facvest	Privada	0,951857	0,854685	1	0,813538
24	1205	Faculdade Educacional Da Lapa	Privada	0,955826	0,959806	0,885178	0,812069
25	375	Universidade Santo Amaro	Privada	0,879217	0,923316	1	0,811795
26	575	Universidade Federal De Minas Gerais	Pública	0,884916	0,912179	1	0,807202
27	18637	Faculdade Imesmercosur	Privada	0,969029	0,888476	0,935754	0,805646
28	1396	Centro Universitário Dinâmica Das Cataratas	Privada	0,940022	0,847459	1	0,79663
29	1510	Centro Universitário Estácio De Santa Catarina	Privada	0,992159	0,93156	0,861588	0,796328
30	570	Universidade Federal Do Rio Grande Do Norte	Pública	0,876564	0,898935	1	0,787974

Fonte: Elaboração própria no R-Studio.

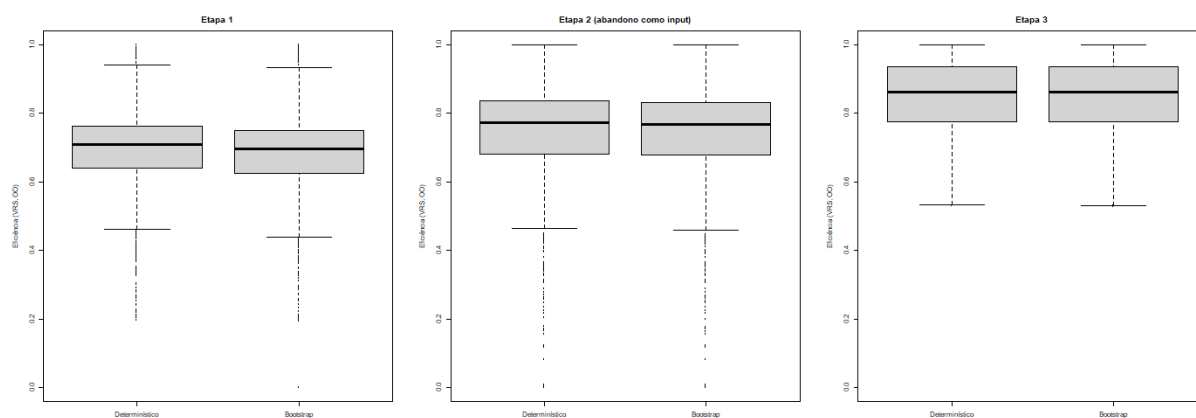
As universidades públicas continuam presentes entre as 30 primeiras, mas com resultados ainda mais ajustados em relação ao modelo clássico. Destacam-se a Universidade de Brasília (UnB), a Universidade Federal do Pará (UFPA), a Universidade Federal do Rio de Janeiro (UFRJ), a Universidade Federal de Minas Gerais (UFMG) e a Universidade Federal do Rio Grande do Norte (UFRN). Todas apresentaram eficiência global acima de 0,78, mantendo-se entre as mais eficientes, ainda que em posições inferiores às ocupadas no modelo determinístico. Esse comportamento demonstra que, mesmo após a correção do viés, as universidades federais de grande porte permanecem relevantes no cenário nacional, reforçando a solidez de sua estrutura acadêmica.

Em síntese, os resultados do NDEA-bootstrap corroboram o padrão identificado no modelo clássico, mas introduzem maior rigor metodológico ao reduzir a incidência de escores máximos e ao oferecer intervalos de confiança mais realistas. Essa abordagem permite avaliar com maior precisão a robustez das estimativas de eficiência e evidencia que, apesar da ampla presença das instituições privadas no topo, universidades públicas tradicionais mantêm posição de destaque e relevância acadêmica no sistema de ensino superior brasileiro.

Para aprofundar a robustez da avaliação de eficiência, aplicou-se o método bootstrap às estimativas obtidas pela Análise Envoltória de Dados (DEA) sob retornos variáveis de escala (VRS). O procedimento permitiu corrigir o viés das medidas determinísticas de eficiência, além de fornecer intervalos de confiança a 95% para cada unidade de decisão (DMU). Assim, tornou-se possível avaliar não apenas a posição relativa das instituições frente à fronteira eficiente, mas também a precisão estatística dessas estimativas.

O Gráfico 5 apresenta os boxplots comparativos entre as eficiências determinísticas e as corrigidas pelo bootstrap para as três etapas do modelo em rede. Observa-se que, após a correção, há uma leve redução central das eficiências e aumento da dispersão, evidenciando o ajuste para eliminar o viés do estimador original.

Gráfico 5 - Boxplots Eficiências corrigidas nas três etapas (2023)

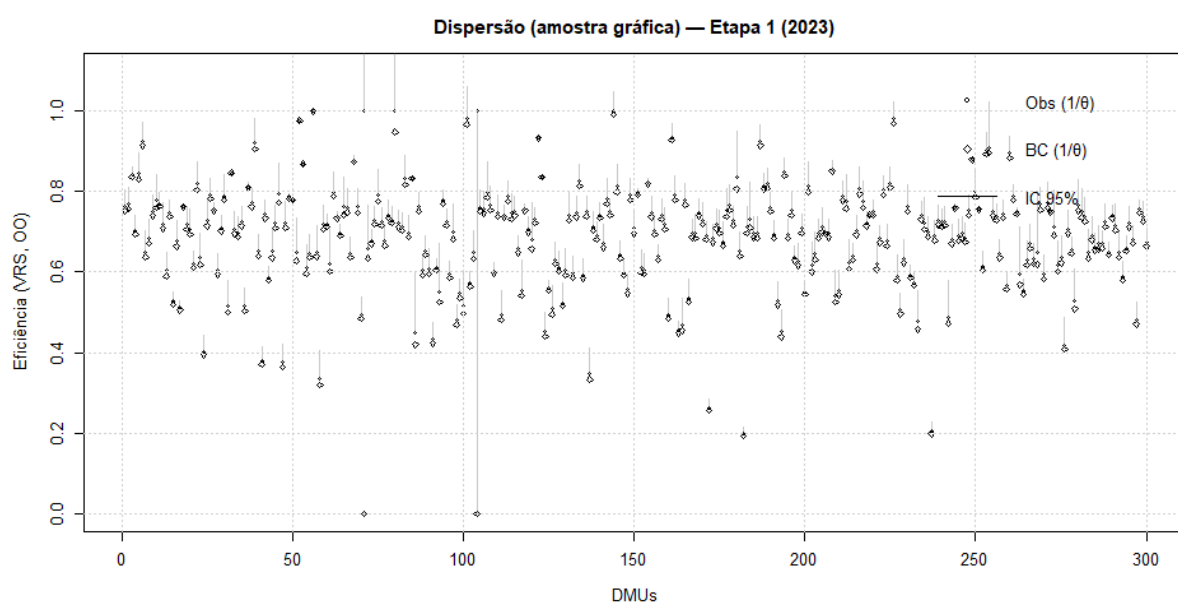


Fonte: Elaboração própria no R-Studio.

Os Gráficos 6, 7 e 8 mostram a dispersão das eficiências individuais das DMUs, já com os respectivos intervalos de confiança, nas etapas 1, 2 e 3, respectivamente. Nota-se que muitas instituições mantêm eficiências consistentes, com intervalos curtos e próximos da fronteira, enquanto outras exibem maior incerteza, refletida em

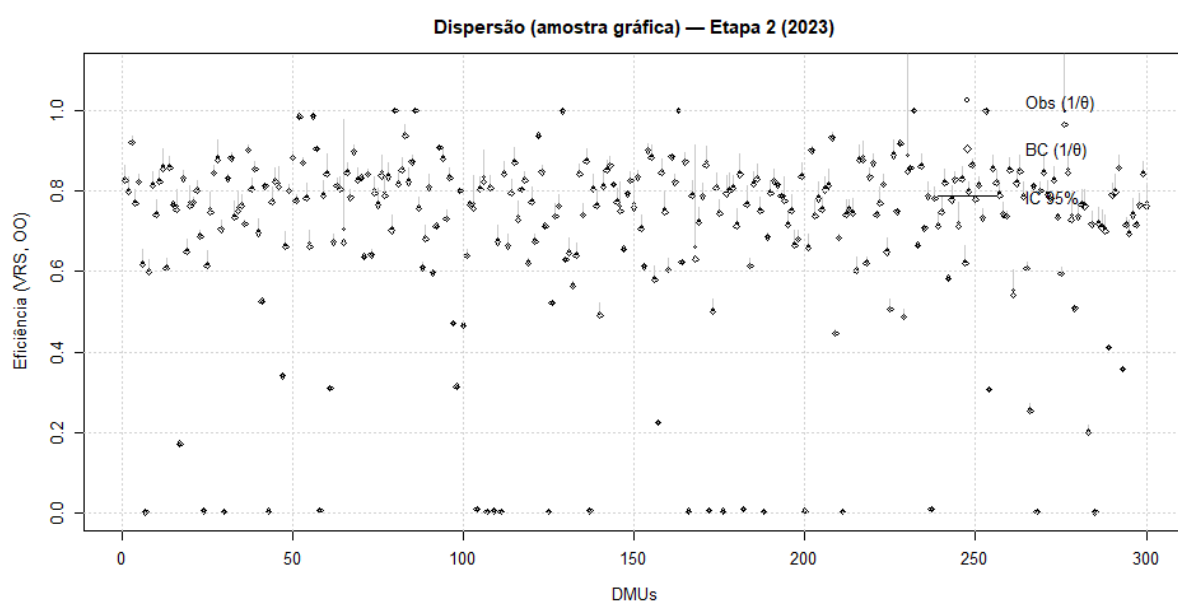
barras mais longas. Ainda com base nesses gráficos, os losangos vazios correspondem ao valor central da eficiência corrigido pelo bootstrap, enquanto os limites representados pelas barras superior e inferior indicam o intervalo de confiança para cada DMU, considerando o viés estimado. Essa representação visual facilita a identificação de instituições com desempenho mais estável em contraste com aquelas que apresentam maior variação.

Gráfico 6 - Dispersão de eficiências com intervalo de confiança (Etapa 1 – 2023)



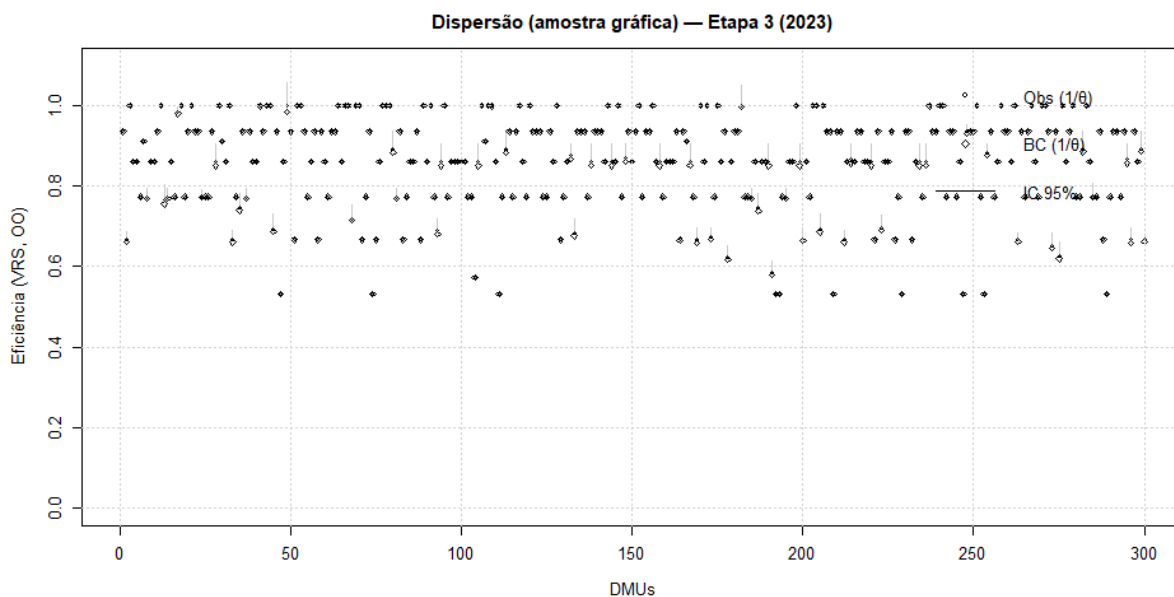
Fonte: Elaboração própria no R-Studio.

Gráfico 7 - Dispersão de eficiências com intervalo de confiança (Etapa 2 – 2023)



Fonte: Elaboração própria no R-Studio.

Gráfico 8 - Dispersão de eficiências com intervalo de confiança (Etapa 3 – 2023)



Fonte: Elaboração própria no R-Studio.

Nota-se, portanto, que os novos índices de eficiência, além de não viesados, apresentam menor variância e se configuram como métricas estatisticamente mais confiáveis. A literatura (Simar; Wilson, 1998; 2007) aponta, contudo, que esse viés poderia ser ainda mais amenizado se fosse elevado o número de reamostragens, por exemplo, de $B = 500$ para $B = 2000$, produzindo intervalos de confiança ainda mais precisos.

Em síntese, a aplicação do bootstrap reforça a credibilidade dos resultados, pois além de fornecer escores corrigidos, acrescenta uma dimensão de confiabilidade estatística que permite interpretações mais seguras e críticas do desempenho institucional.

4.5.2 Resultados de Eficiência entre Redes de Ensino Superior

No presente estudo, o foco recai sobre a análise comparativa da eficiência técnica aplicada ao processo operacional das Instituições de Ensino Superior (IES) brasileiras, considerando dois segmentos complementares: a) as IES públicas federais, financiadas por recursos orçamentários diretos da União, e b) as IES privadas, que podem receber alunos beneficiários do Fies, instituído pela Lei n. 10.260, de 12 de julho de 2001 (Pontuschka, 2016).

Embora não se trate de uma análise da eficiência do gasto público, a apresentação dos valores médios de dispêndios diretos e indiretos entre 2019 e 2023

auxilia na contextualização do porte financeiro de cada modalidade de oferta. A Tabela 11 sintetiza esses valores, obtidos no Portal da Transparência da União (Brasil, [s.d.]b; [s.d.]c), considerando as despesas com ensino superior (função 12, subfunção 364) e os desembolsos vinculados ao programa Fies.

Tabela 11 - Gastos Direto e Indireto, 2019–2023

Ano	Gasto Direto (R\$ bi)	Gasto Indireto (Fies) (R\$ bi)
2019	30,41	8,95
2020	29,11	5,09
2021	28,81	4,39
2022	30,43	3,99
2023	33,88	4,04
Média	30,50	5,71

Fonte: Elaboração própria com dados de Brasil ([s.d.]b, [s.d.]c)

Entre 2019 e 2023, os gastos diretos com as IES públicas federais permaneceram relativamente estáveis — variando de R\$ 28,81 bi a R\$ 33,88 bi (média R\$ 30,50 bi). Já os desembolsos indiretos via Fies apresentaram forte retração, caindo de R\$ 8,95 bi (2019) para R\$ 4,04 bi (2023), uma redução de 54,8% no período (média R\$ 5,71 bi). A participação do Fies no gasto total (direto + indireto) também recuou de 22,7% em 2019 para 10,7% em 2023, evidenciando a perda de peso relativo do financiamento indireto no arranjo de gastos do ensino superior federal.

A Tabela 12 apresenta as estatísticas descritivas dos escores de eficiência técnica estimados por meio da DEA, considerando retornos variáveis de escala (VRS) e orientação a outputs, para o período de 2019 a 2023. Os escores reportados decorrem da aplicação do modelo DEA no primeiro estágio da análise empírica, no qual se avalia a eficiência relativa das IES a partir da relação entre os insumos e produtos definidos para cada etapa do processo produtivo.

Tabela 12 - Estatísticas descritivas das eficiências por rede e ano (2019–2023)

Rede	Ano	Média	Mediana	Desvio-padrão	N
Privada	2019	0,396	0,443	0,218	1794
Privada	2020	0,391	0,422	0,198	1794
Privada	2021	0,397	0,425	0,193	1794
Privada	2022	0,402	0,423	0,175	1794
Privada	2023	0,418	0,428	0,168	1794
Pública	2019	0,606	0,593	0,137	109
Pública	2020	0,574	0,562	0,145	109
Pública	2021	0,573	0,561	0,127	109

Rede	Ano	Média	Mediana	Desvio-padrão	N
Pública	2022	0,573	0,561	0,126	109
Pública	2023	0,586	0,588	0,137	109

Fonte: Elaboração própria no R-Studio.

Os resultados mostram que, entre 2019 e 2023, as IES públicas mantêm níveis de eficiência sistematicamente superiores aos das privadas, com médias entre 0,57 e 0,61, ligeira redução entre 2020 e 2022 e recuperação em 2023 (0,586). As instituições privadas partem de um patamar médio inferior (0,396 em 2019), mas apresentam trajetória ascendente e consistente, alcançando 0,418 em 2023.

O comportamento do desvio-padrão sugere uma redução gradual do diferencial médio de eficiência entre os dois segmentos ao longo do período analisado, que passa de aproximadamente 0,218 em 2019 para 0,168 em 2023, correspondendo a uma diminuição em torno de 20%. Tal convergência decorre, sobretudo, do ganho incremental de eficiência observado entre as IES privadas, enquanto as instituições públicas mantêm níveis relativamente estáveis de desempenho. Essa dinâmica indica um processo de aproximação entre os segmentos, ainda que os níveis médios de eficiência das IES públicas permaneçam superiores em todo o período.

A análise da dispersão reforça esse diagnóstico. As IES privadas demonstram convergência de desempenho — o desvio-padrão cai de 0,218 para 0,168 —, o que indica maior homogeneidade institucional ao longo do tempo. Nas públicas, a variabilidade permanece estável, refletindo processos administrativos mais uniformes. Em ambos os grupos, entretanto, a média e a mediana inferiores a 1 revelam que ainda há amplo espaço para ganhos de eficiência.

Essas evidências devem ser interpretadas à luz da estrutura amostral — o número de observações é muito superior nas instituições privadas (N=1.776) —, mas o padrão geral é consistente: as IES públicas permanecem mais eficientes em média, enquanto as privadas mostram trajetória de aprimoramento, reduzindo heterogeneidades internas.

A partir desses resultados, o estudo avança para examinar se a participação no Fies constitui um fator associado à melhoria da eficiência técnica aplicada ao processo operacional das IES privadas, hipótese que será verificada nos modelos econométricos e de aprendizado de máquina no segundo estágio da análise.

4.6 FATORES EXÓGENOS

Após a mensuração da eficiência global das IES, tanto pela abordagem determinística quanto pela aplicação do bootstrap, esta seção dedica-se à análise dos fatores exógenos associados às diferenças observadas nos escores de eficiência. O objetivo é identificar quais características socioeconômicas, institucionais e locais se mostram mais relevantes para explicar o desempenho relativo das instituições e como esses resultados dialogam com a literatura empírica existente.

A análise foi conduzida por meio de algoritmos de Machine Learning (Elastic Net, Random Forest e XGBoost), tendo como variável dependente a Eficiência Global livre de viés, restrita ao intervalo (0,1). As variáveis explicativas contemplaram dimensões estruturais e contextuais do ambiente das Instituições de Ensino Superior (IES), permitindo capturar diferentes condicionantes externos ao processo produtivo educacional.

A comparação entre os algoritmos revela diferenças claras quanto à capacidade preditiva, medida pelo erro quadrático médio (RMSE).

Tabela 13 - Resultados dos modelos de aprendizado de máquina

Modelo	Estrutura de Saída	Principais Variáveis (mais relevantes)	RMSE Médio	Interpretação
Elastic Net	Fórmula linear explícita com coeficientes (β)	PIBpc (+), Densidade (+/-), Ano_mec (+), Dummy_FIES (+), Lat/Lon (efeitos regionais)	0,164	Fornecer direção e magnitude das associações, mas com menor capacidade preditiva.
Random Forest	Ensemble de árvores, sem fórmula explícita	Ano_mec, Dummy_FIES, Densidade, Lat/Lon (maior importância); PIBpc (menor impacto)	0,152	Captura relações não lineares e interações; indica efeito marginal positivo do Dummy_FIES.
XGBoost	Gradient Boosting, sem fórmula explícita	Ano_mec, Dummy_FIES, Densidade, Lat/Lon, PIBpc	0,147	Melhor capacidade preditiva; confirma o peso da maturidade institucional e da influência do Fies como fator contextual.

Fonte: Elaboração própria a partir dos resultados do ML.

Conforme apresentado na Tabela 13, o XGBoost apresentou o melhor desempenho (RMSE \approx 0,149), seguido pelo Random Forest (RMSE \approx 0,154) e pelo Elastic Net (RMSE \approx 0,168). Esses resultados indicam que abordagens capazes de capturar relações não lineares e interações complexas oferecem vantagens na explicação da eficiência institucional, corroborando evidências da literatura recente

sobre aplicações de Machine Learning em análises educacionais e organizacionais (Janiesch; Zschech; Heinrich, 2021; Bi *et al.*, 2019; Sarker, 2021).

Os resultados do Elastic Net permitem identificar relações diretas entre variáveis ambientais e eficiência, ainda que com menor capacidade preditiva. Observou-se associação positiva entre o PIB per capita municipal e a eficiência, em linha com os achados de Arbula Blecich (2024), que destacam o papel do contexto econômico regional na eficiência educacional. A densidade demográfica apresentou efeito de menor magnitude, sugerindo comportamento ambivalente, conforme já indicado por Afonso *et al.* (2005), que apontam que a concentração populacional pode tanto favorecer economias de escala quanto gerar pressões sobre a infraestrutura.

A idade institucional (ANO_MEC) mostrou-se positivamente associada à eficiência, reforçando evidências de que instituições mais consolidadas tendem a apresentar melhor desempenho operacional, em consonância com Siqueira (2015) e com os achados mais recentes de Chen *et al.* (2025) sobre maturidade organizacional. As variáveis geoespaciais (LAT/LON) apresentaram relevância secundária nesse modelo. Apesar da clareza interpretativa, o desempenho inferior do Elastic Net (RMSE $\approx 0,168$) sugere limitações na captura de relações não lineares, conforme já apontado na literatura comparativa de métodos (Kuhn e Johnson, 2013).

No Random Forest, a análise da importância relativa das variáveis indicou maior peso para o PIB per capita municipal e a densidade demográfica, seguidos pelos fatores espaciais (LAT/LON), enquanto a idade institucional apresentou contribuição marginal. Esses resultados reforçam os achados de Afonso *et al.* (2005) e de Nazarko e Šaparauskas (2014), que destacam a relevância do contexto socioeconômico e territorial na explicação da eficiência organizacional.

O desempenho intermediário do Random Forest (RMSE $\approx 0,154$) confirma sua robustez na identificação de padrões não lineares, conforme destacado por Breiman (2001), posicionando-o como alternativa que equilibra capacidade explicativa e interpretabilidade.

O XGBoost apresentou os resultados mais consistentes do ponto de vista preditivo (RMSE $\approx 0,149$). A análise de importância das variáveis revelou a idade institucional (ANO_MEC) como principal determinante da eficiência global, seguida pela densidade demográfica e pelos fatores espaciais (LAT/LON). Em contraste com os modelos anteriores, o PIB per capita municipal apresentou efeito marginal.

Esse resultado sugere que a eficiência das IES está fortemente condicionada por fatores ligados à trajetória histórica e à maturidade institucional, mais do que exclusivamente ao nível de desenvolvimento econômico local. Tal evidência amplia os achados de Afonso *et al.* (2005), ao indicar que a influência socioeconômica pode ser mediada por fatores institucionais e espaciais, e converge com os resultados de Santos, Silva Júnior e Nunes (2022), que identificam maior eficiência média em instituições com estruturas organizacionais mais consolidadas.

Embora o XGBoost não forneça uma equação explícita, sua robustez empírica confirma sua utilidade em contextos caracterizados por elevada complexidade estrutural, conforme argumentado por Chen e Guestrin (2016).

Em síntese, a comparação entre os modelos evidencia sua complementaridade analítica. O Elastic Net oferece uma visão interpretativa clássica, baseada em coeficientes que permitem discutir o sinal e a magnitude das associações; o Random Forest amplia a robustez ao identificar determinantes-chave mesmo em cenários com multicolinearidade; e o XGBoost representa a fronteira tecnológica ao maximizar o poder preditivo e capturar relações complexas, ainda que sem a possibilidade de equação fechada. Assim, a estratégia combinada adotada mostra-se metodologicamente sólida, ao conciliar explicação econômica e capacidade preditiva.

O processo de tuning revelou diferenças claras de desempenho entre os modelos. O XGBoost apresentou a menor raiz quadrada do erro quadrático médio (RMSE \approx 0,149), seguido pelo Random Forest (RMSE \approx 0,154) e pelo Elastic Net (RMSE \approx 0,168). O RMSE mede a distância média, em termos quadráticos, entre valores observados e previstos e quanto menor, melhor o ajuste preditivo do modelo.

No XGBoost, a análise da importância relativa das variáveis evidenciou a idade institucional (*ano_mec*) como principal determinante da eficiência, seguida pela densidade demográfica e pela localização geográfica (latitude e longitude), ficando o PIB per capita municipal em posição secundária. Esses achados reforçam que a trajetória histórica das instituições e seu contexto espacial exercem maior influência que indicadores econômicos isolados.

Assim, o XGBoost foi selecionado como modelo final ajustado. Observa-se que o XGBoost apresentou o melhor resultado em termos de predição (menor RMSE), enquanto o Elastic Net, embora menos acurado, forneceu coeficientes interpretáveis em formato clássico, reforçando a complementaridade entre os modelos.

Os resultados obtidos consolidaram o XGBoost como a abordagem mais robusta para a análise, dada sua capacidade de lidar com relações não lineares e interações complexas entre as variáveis. O modelo reforçou que a eficiência das IES é condicionada por múltiplas dimensões externas e internas, evidenciando a utilidade de métodos de aprendizado de máquina para ampliar a compreensão do fenômeno além das abordagens econométricas tradicionais. Esse resultado não apenas sustenta a escolha metodológica realizada, mas também abre espaço para futuras investigações que explorem novas variáveis e contextos comparativos.

Tabela 14 - Importância das variáveis ambientais – XGBoost

Variável	Importância
ano_mec	0,842519
dummy_Fies	0,065184
dens_mun	0,039872
tp_rede	0,027406
lat	0,012497
lon	0,008923
in_capital	0,002804
pibpc_mun	0,001895

Fonte: Elaboração própria a partir dos resultados do ML.

A análise da importância relativa das variáveis no modelo final (XGBoost) evidencia de forma clara a predominância da idade institucional (ano_mec), que apresentou peso de 0,842, muito acima dos demais preditores (Tabela 14). Esse resultado confirma a hipótese levantada em estudos anteriores, como os de Siqueira (2015) e Santos *et al.* (2022), que destacam a maturidade institucional como fator decisivo para a eficiência, dada sua associação com processos de gestão mais consolidados, acúmulo de capital humano e experiência administrativa ao longo do tempo.

Em segundo lugar, destaca-se a variável dummy_Fies (0,065), indicando que a participação das instituições no programa de financiamento estudantil exerce influência positiva sobre a eficiência global. Esse resultado sugere que o Fies atua como um mecanismo de estabilidade financeira, sobretudo no setor privado, contribuindo para a manutenção da capacidade operacional e para a melhoria relativa da produtividade, mesmo em contextos de restrição orçamentária. A evidência reforça a literatura que relaciona o acesso a recursos previsíveis e o financiamento estudantil à sustentabilidade das IES e à expansão do acesso ao ensino superior.

A densidade demográfica municipal (*dens_mun*) manteve relevância intermediária (0,040), corroborando parcialmente os achados de pesquisas que associam a concentração populacional a maior pressão sobre os serviços educacionais e maior demanda por ensino superior. Embora sua importância seja inferior à da idade institucional, o resultado alinha-se aos estudos de Arbula Blecich (2024), que indicam que contextos urbanos mais densos oferecem tanto desafios quanto oportunidades adicionais para as IES, afetando sua eficiência técnica aplicada ao processo operacional.

O tipo de rede (*tp_rede*) aparece em seguida (0,027), confirmando a literatura que diferencia o desempenho entre IES públicas e privadas. Pesquisas como as de Santos *et al.* (2022) já apontavam que diferenças na governança, nas fontes de financiamento e nas estratégias institucionais impactam diretamente os níveis de eficiência, e os resultados do presente estudo reforçam essa distinção, ainda que parte dessa diferença seja absorvida pela variável *dummy_Fies*.

Esse resultado é coerente com as evidências obtidas na etapa explicativa baseada na regressão Beta, apresentada na Seção 4.7 EFEITOS DO FIES SOBRE A EFICIÊNCIA DAS IES PRIVADAS, na qual o tipo de rede (*tp_rede*) apresentou coeficiente negativo e estatisticamente significativo ($\beta = -1,221$; $p < 0,001$), indicando que, em média, as Instituições de Ensino Superior privadas exibem níveis de eficiência inferiores aos das instituições públicas. Contudo, a inclusão da variável *dummy_Fies* no modelo revelou que parte substancial dessa diferença é mitigada quando as IES privadas participam do programa de financiamento estudantil.

Enquanto o efeito negativo associado à condição privada permanece relevante, o coeficiente positivo e robusto da *dummy_Fies* ($\beta = 1,135$; $p < 0,001$) evidencia que o Fies atua como fator compensatório, elevando a eficiência relativa das instituições privadas beneficiadas e aproximando seus desempenhos dos observados no setor público. Dessa forma, tanto a análise baseada em Machine Learning quanto a regressão Beta convergem ao indicar que as diferenças de eficiência entre redes não decorrem exclusivamente da natureza pública ou privada das IES, mas são fortemente condicionadas pelos mecanismos de financiamento e pelas condições institucionais associadas à participação no Fies.

Por outro lado, as variáveis geográficas — latitude (0,012) e longitude (0,009) — apresentaram importância reduzida. Esse achado dialoga com trabalhos anteriores que utilizam indicadores espaciais como proxies de localização institucional, indicando

que a dimensão territorial tende a exercer influência indireta sobre o desempenho das instituições, frequentemente mediada por fatores socioeconômicos e estruturais (Lee, 2011; Martínez-Campillo; Fernandez-Santos, 2020; Agasisti *et al.*, 2021). De forma semelhante, a variável *in_capital* (0,003) também se mostrou de baixa influência, contrastando com parte da literatura que associa a localização em centros urbanos ou regiões economicamente dinâmicas a potenciais vantagens institucionais (Guironnet; Peypoch, 2018). Esses resultados sugerem que a localização em capitais estaduais, isoladamente, não garante ganhos significativos de eficiência, a menos que esteja associada a condições institucionais, econômicas e organizacionais favoráveis.

Por fim, o PIB per capita municipal (*pibpc_mun*) apresentou a menor relevância (0,002), divergindo parcialmente de estudos que apontavam a prosperidade econômica do entorno como fator de apoio ao desempenho institucional. Tal resultado sugere que o nível de riqueza local não se traduz diretamente em ganhos de eficiência, sendo possivelmente absorvido por variáveis mais próximas da realidade operacional, como o perfil socioeconômico dos estudantes e as condições institucionais internas.

À luz dessas evidências, torna-se necessário aprofundar a análise sobre o papel específico dos instrumentos de financiamento estudantil na explicação das diferenças de desempenho entre as instituições, em especial no setor privado. Nesse sentido, a seção seguinte dedica-se a examinar de forma explícita os efeitos do Fies sobre a eficiência técnica aplicada ao processo operacional das Instituições de Ensino Superior privadas, por meio de uma abordagem estatística complementar baseada em regressão Beta.

4.7 EFEITOS DO FIES SOBRE A EFICIÊNCIA DAS IES PRIVADAS

O uso da regressão Beta para analisar os efeitos de variáveis exógenas sobre os escores de eficiência obtidos pela DEA é justificado pelas características estatísticas desses escores, que se encontram restritos ao intervalo (0,1), apresentam distribuição assimétrica, heterocedasticidade e violação da normalidade dos resíduos, o que inviabiliza a aplicação de métodos tradicionais de regressão linear, como o estimador de Mínimos Quadrados Ordinários (OLS). A regressão Beta mostra-se particularmente adequada nesse contexto por respeitar o suporte da variável dependente, acomodar diferentes formatos de distribuição e permitir interpretação econômica direta dos efeitos marginais das variáveis exógenas sobre a eficiência

média condicional. Além disso, a adoção da regressão Beta constitui o eixo inferencial da análise, fornecendo estimativas estatisticamente interpretáveis sob um arcabouço paramétrico bem definido. As técnicas de aprendizado de máquina, por sua vez, são empregadas de forma complementar à regressão clássica, sobretudo para lidar com problemas recorrentes em análises multivariadas, como multicolinearidade entre variáveis explicativas, relações não lineares e interações complexas, bem como para avaliar o desempenho preditivo dos modelos. Dessa forma, a integração entre regressão Beta e aprendizado de máquina não configura abordagens concorrentes, mas estratégias analíticas complementares, nas quais a regressão assegura validade inferencial e interpretação causal aproximada, enquanto o ML amplia a capacidade explicativa e preditiva da análise dos determinantes da eficiência institucional.

A análise por meio da regressão Beta não se restringe à comparação entre tipos de instituições, mas busca avaliar de forma conjunta e condicional os efeitos de diferentes variáveis ambientais sobre os níveis de eficiência técnica aplicada ao processo operacional das IES. Nesse contexto, a variável *dummy_Fies* foi incorporada com o objetivo de verificar se a participação das IES privadas no Fies está associada a variações significativas nos escores de eficiência, após o controle por características socioeconômicas, institucionais e locais. A regressão Beta (Ferrari; Cribari-Neto, 2004) permite a estimação do impacto marginal dessas variáveis sobre uma variável dependente contínua restrita ao intervalo unitário, como os escores de eficiência derivados da DEA. Conforme destacado por Simar e Wilson (2007), trata-se de uma alternativa estatisticamente consistente para o segundo estágio em modelos não paramétricos, viabilizando a interpretação isolada dos efeitos do ambiente operacional sobre o desempenho institucional.

Tabela 15 - Resultados da Regressão Beta para a Eficiência Global (EG)

Variável	Coefficient e (β)	Erro-padrão	Valor z	p-valor	Significância	Interpretação
(Intercepto)	73.55	0.152	482.902	< 2e-16	***	Constante do modelo.
<i>dens_mun</i>	0.0000607	0.0000106	5.729	1.01e-08	***	Associação positiva entre densidade demográfica municipal e eficiência.
<i>piibpc_mun</i>	0.0000021	0.0000026	0.830	0.407	–	Variável não apresentou significância estatística.
<i>ano_mec</i>	-0.03743	0.0000378	-990.783	< 2e-16	***	Ano de credenciamento associado negativamente ao indicador de eficiência.
<i>lat</i>	-0.01136	0.00209	-5.436	5.45e-08	***	Localização geográfica (latitude) apresenta

Variável	Coefficiente (β)	Erro-padrão	Valor z	p-valor	Significância	Interpretação
lon	-0.01581	0.00267	-5.922	3.18e-09	***	Localização geográfica (longitude) apresenta associação negativa com a eficiência.
tp_rede	-1.221	0.0723	-16.884	< 2e-16	***	Diferença estatisticamente significativa entre redes pública e privada.
in_capital	0.1543	0.0353	4.367	1.26e-05	***	Instituições localizadas em capitais apresentam associação positiva com eficiência.
dummy_Fies	1.135	0.0371	30.577	< 2e-16	***	Participação no Fies associada positivamente ao indicador de eficiência.

Fonte: Elaboração própria a partir dos resultados da Regressão Beta.

Nota 1: *** $p < 0,01$; ** $p < 0,05$; * $p < 0,10$.

Nota 2: A tabela apresenta o conjunto de Instituições de Ensino Superior (IES) consideradas na análise empírica do estudo, incluindo instituições públicas federais e instituições privadas participantes do Fies no período analisado. Não estão incluídas instituições estaduais, municipais ou privadas que não participam do programa. A seleção das instituições foi realizada a partir das bases de dados do Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep) e das informações disponíveis sobre participação no Fies, conforme descrito nos procedimentos metodológicos do estudo.

Além da análise dos coeficientes estimados, o desempenho do modelo de regressão Beta foi avaliado por meio da raiz do erro quadrático médio (Root Mean Squared Error – RMSE), métrica que sintetiza a magnitude média dos desvios entre os valores observados e os valores ajustados. Valores menores de RMSE indicam maior capacidade preditiva e melhor ajuste global do modelo. Os resultados mostram que, embora a regressão Beta apresente desempenho consistente e adequado à natureza limitada dos escores de eficiência, os modelos de aprendizado de máquina apresentam valores de RMSE inferiores, refletindo maior acurácia preditiva. Essa diferença decorre da capacidade dos algoritmos de ML em capturar relações não lineares, interações complexas e heterogeneidade entre as IES, aspectos que tendem a ser parcialmente restringidos em modelos paramétricos. Assim, enquanto a regressão Beta se destaca pela interpretabilidade econômica e pela inferência estatística, os modelos de aprendizado de máquina apresentam superioridade em termos de desempenho preditivo, configurando abordagens complementares na análise explicativa da eficiência institucional.

À luz do desempenho preditivo e inferencial dos modelos, os resultados combinados da regressão Beta e das técnicas de aprendizado de máquina indicam uma associação positiva e estatisticamente significativa entre a participação no Fies e os escores de eficiência das IES privadas. O coeficiente estimado para a variável *dummy_Fies* ($\beta = 1,135$; $p < 0,001$) confirma que as instituições participantes do programa tendem a apresentar níveis mais elevados de eficiência técnica em comparação às não participantes. Esse achado reforça a literatura que aponta os mecanismos de financiamento estudantil como instrumentos de fortalecimento institucional e de estabilidade financeira (Hoff, 2007).

Do ponto de vista técnico, o modelo de regressão Beta evidenciou ainda que a densidade demográfica (*dens_mun*) possui relação positiva e estatisticamente significativa com a eficiência, sugerindo que contextos urbanos mais densos favorecem melhores resultados operacionais. Em contrapartida, o PIB per capita municipal (*pibpc_mun*) não apresentou significância estatística ($p = 0,407$), indicando que a prosperidade econômica regional não se traduz, necessariamente, em maior eficiência das IES.

O ano de credenciamento (*ano_mec*) apresentou efeito negativo expressivo ($\beta = -0,037$; $p < 0,001$), sugerindo que instituições mais antigas podem enfrentar estruturas administrativas mais rígidas e menor flexibilidade operacional. As variáveis espaciais latitude (*lat*) e longitude (*lon*) também exibiram coeficientes negativos e estatisticamente significativos, evidenciando a persistência de desigualdades regionais na eficiência das Instituições de Ensino Superior. O sinal negativo da latitude indica que, *ceteris paribus*, IES localizadas em regiões mais ao Norte tendem a apresentar menores níveis de eficiência, enquanto o coeficiente negativo da longitude sugere desempenho inferior das instituições situadas mais a Leste, mantidas constantes as demais variáveis do modelo.

Em contraste, a variável *in_capital* mostrou associação positiva e significativa com a eficiência ($\beta = 0,154$; $p < 0,001$), denotando a existência de vantagem locacional para IES sediadas em capitais estaduais. Já o tipo de rede (*tp_rede*) apresentou sinal negativo, confirmando que as IES privadas, em geral, exibem eficiência média inferior às públicas, embora as beneficiadas pelo Fies se destaquem positivamente dentro do grupo privado.

No que se refere à inclusão das Instituições de Ensino Superior públicas no modelo, estas foram consideradas como grupo de referência (baseline) na

especificação da regressão. A variável categórica *tp_rede* foi codificada de forma binária, distinguindo instituições públicas e privadas, enquanto a variável *dummy_Fies* foi definida apenas para as instituições privadas participantes do programa. Dessa forma, as universidades públicas federais permaneceram no modelo como parte integrante da amostra e como parâmetro de comparação para os coeficientes estimados. Essa estratégia permite identificar diferenças relativas de eficiência entre os dois segmentos institucionais, ao mesmo tempo em que possibilita estimar o efeito específico da participação no Fies dentro do conjunto das instituições privadas. Assim, as instituições públicas não foram excluídas da análise; ao contrário, desempenham papel fundamental na estrutura do modelo, funcionando como categoria de referência para a interpretação dos coeficientes associados às demais variáveis institucionais.

O efeito positivo e robusto da variável *dummy_Fies* reforça que o programa exerce influência direta sobre a eficiência das IES privadas, atuando como um mecanismo de estabilidade financeira que reduz a inadimplência e amplia a previsibilidade de receitas, viabilizando investimentos em infraestrutura, tecnologia e qualificação docente. Essa estabilidade repercute em melhorias nos processos administrativos e na qualidade dos resultados operacionais, o que explica a convergência gradual dos níveis médios de eficiência das IES privadas em direção às públicas. A integração entre abordagens paramétricas (regressão Beta) e não paramétricas (DEA e ML) amplia a robustez das inferências, permitindo uma compreensão mais abrangente das fontes de variação da eficiência institucional — prática metodológica recomendada em análises de múltiplos estágios (Tone; Tsutsui, 2014).

Contudo, os efeitos não se mantêm uniformes ao longo do tempo. Em períodos de redução de contratos do Fies ou de restrição orçamentária, observa-se queda na eficiência das IES mais dependentes do programa, evidenciando a sensibilidade dessas instituições à continuidade das políticas de financiamento estudantil. Essa volatilidade reforça a importância de estabilidade regulatória e previsibilidade fiscal para assegurar o equilíbrio entre a expansão do acesso e a sustentabilidade financeira do sistema.

Em síntese, os resultados confirmam que o Fies atua como um mecanismo complementar de equilíbrio e eficiência no sistema de ensino superior brasileiro, ao fortalecer financeiramente as IES privadas e reduzir disparidades operacionais em relação às públicas. Entretanto, sua efetividade de longo prazo depende de políticas

estáveis e de uma gestão institucional madura, capaz de converter o suporte financeiro em ganhos permanentes de produtividade, qualidade e equidade educacional.

4.8 AVALIAÇÃO DINÂMICA E ÍNDICE DE PRODUTIVIDADE DE MALMQUIST

A análise da produtividade das IES foi realizada por meio do Índice de Malmquist (IM), calculado em cada uma das três etapas do modelo em rede, de modo a identificar em quais fases do processo produtivo — captação, eficiência interna e eficiência externa — ocorreram avanços ou retrocessos ao longo do período de 2019 a 2023. Essa escolha metodológica permite superar a visão de caixa-preta, oferecendo uma compreensão mais detalhada sobre os mecanismos de geração de resultados em diferentes dimensões da atividade acadêmica.

O Índice de Produtividade de Malmquist (IPM) é decomposto em três componentes que fornecem informações complementares. O primeiro deles é a mudança na eficiência técnica pura (PECH), que capta a variação na capacidade das instituições de utilizar seus recursos de forma eficiente. O segundo é a mudança tecnológica (TC), que mede os deslocamentos da fronteira de produção e expressa a incorporação de inovações. O terceiro é a mudança na eficiência de escala (SECH), que avalia a adequação do porte das instituições ao nível ótimo de operação. Em qualquer um desses componentes, valores superiores a 1 indicam ganhos de produtividade, enquanto valores inferiores a 1 evidenciam retrocessos.

O componente PECH reflete a evolução da capacidade gerencial e operacional das IES, desconsiderando os efeitos de escala. Valores superiores a 1 indicam avanços na gestão interna e nos processos acadêmicos, enquanto valores abaixo de 1 sinalizam dificuldades em acompanhar os padrões de eficiência observados entre as instituições.

A mudança tecnológica (TC) captura deslocamentos da fronteira de produção, apontando inovações ou retrocessos tecnológicos que impactam o conjunto das instituições. Valores acima de 1 sugerem progresso tecnológico, enquanto valores inferiores a 1 indicam perda de dinamismo ou ausência de investimentos.

A eficiência de escala (SECH) mostra se as IES estão operando em dimensão ótima de produção. Valores superiores a 1 refletem ganhos de escala, ao passo que valores inferiores a 1 sugerem subutilização da estrutura ou desperdício de recursos.

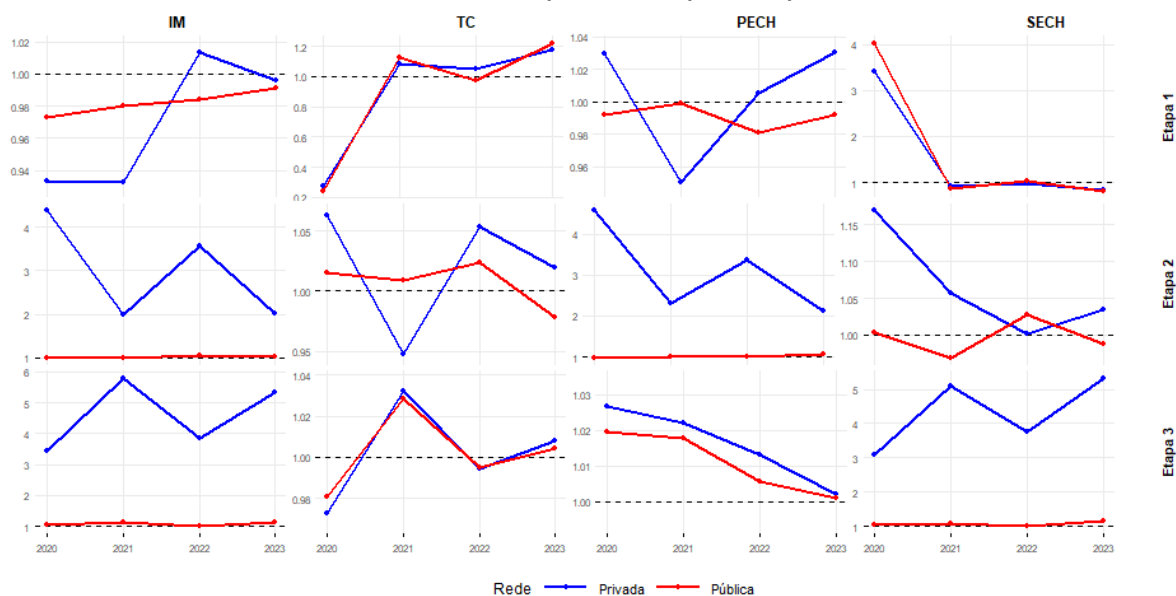
Na etapa de eficiência de captação, os valores do IM mantiveram-se próximos da unidade, sugerindo relativa estabilidade na relação entre recursos humanos, técnicos e infraestrutura frente às matrículas. O componente TC apresentou maior volatilidade, sinalizando diferentes ritmos de incorporação de inovações no processo de atração de alunos. O desempenho nessa etapa sugere que, embora não tenha havido grandes saltos de produtividade, também não se verificaram perdas significativas, o que caracteriza um comportamento de manutenção.

Na etapa de eficiência interna, que relaciona matrículas com concluintes e abandono, os índices de IM foram mais elevados, sobretudo em 2020 e 2022. Esse resultado indica que houve ganhos expressivos na transformação das matrículas em diplomados, ainda que influenciados por picos de SECH. A melhora nesta etapa pode estar associada tanto a estratégias institucionais voltadas à permanência estudantil quanto a ajustes de escala que potencializaram a utilização dos recursos disponíveis.

Na etapa de eficiência externa, responsável por medir os resultados finais a partir da avaliação dos cursos e indicadores de qualidade, observaram-se maiores variações. Em 2021 e 2022, os índices apresentaram patamares mais altos, sustentados por ganhos de escala e de eficiência técnica, mas também acompanhados de oscilações que indicam heterogeneidade no desempenho entre as instituições. Esse comportamento revela que a etapa final da cadeia produtiva do ensino superior é a mais sujeita a variações de contexto, refletindo diferenças institucionais na qualidade dos cursos ofertados e na capacidade de gerar impacto social.

O Gráfico 9 apresenta a evolução do índice global de produtividade de Malmquist (IM) e de seus componentes, desagregados por etapas do processo produtivo educacional e por tipo de rede institucional, no período de 2019 a 2023.

Gráfico 9 - Médias anuais do IM e componentes por etapas, 2019–2023



Fonte: Elaboração própria

A leitura inicial do Gráfico 9 revela que as trajetórias de produtividade não são homogêneas ao longo do tempo nem entre as redes de ensino, tampouco entre as diferentes etapas do processo produtivo educacional, evidenciando a importância da decomposição do Índice de Produtividade de Malmquist para a compreensão da dinâmica produtiva das Instituições de Ensino Superior (IES). Os resultados indicam que os movimentos de ganhos e perdas de produtividade se distribuem de forma diferenciada entre etapas, componentes e tipos de rede, afastando interpretações agregadas ou lineares do desempenho institucional.

De forma geral, observa-se que as IES privadas exibem maior variabilidade intertemporal dos índices, sobretudo nas Etapas 2 e 3, nas quais se concentram oscilações mais pronunciadas tanto no índice global (IM) quanto nos componentes de eficiência de escala (SECH) e de eficiência técnica pura (PECH). Esse comportamento sugere maior sensibilidade dessas instituições a mudanças no ambiente econômico e institucional, resultado que corrobora parcialmente evidências reportadas por estudos anteriores que associam maior dinamismo das IES privadas a maior exposição a choques externos, ajustes organizacionais frequentes e reconfigurações estratégicas (Agasisti *et al.*, 2021; Martínez-Campillo; Fernández-Santos, 2020).

A análise dos componentes indica que a mudança tecnológica (TC) é o elemento com maior variabilidade visual ao longo do período analisado,

especialmente entre as IES privadas e nas etapas intermediárias do processo produtivo. Tal evidência é coerente com a literatura internacional, que aponta o componente TC como o mais sensível a alterações no ambiente regulatório, tecnológico e econômico (Parteka; Wolszczak-Derlacz, 2013). No entanto, diferentemente do que alguns estudos sugerem, não se observa um padrão claro e persistente de ganhos tecnológicos acumulativos ao longo do período, o que pode indicar dificuldades de sustentação desses avanços no médio prazo.

A interpretação desses movimentos pode ser enriquecida pela consideração de fatos temporais relevantes ocorridos no período analisado. Em particular, os anos de 2020 e 2021 coincidem com o choque exógeno provocado pela pandemia da COVID-19, que exigiu das IES uma reorganização abrupta de seus processos acadêmicos e administrativos, com migração emergencial para o ensino remoto e posterior adoção de arranjos híbridos. A literatura aponta que choques dessa natureza tendem a se manifestar principalmente por meio do componente de mudança tecnológica e de ajustes de escala, o que torna compatíveis com esse contexto as oscilações observadas nos componentes TC e SECH, especialmente entre as IES privadas. Ademais, o período analisado também se insere em um contexto de reconfiguração das políticas de financiamento estudantil, marcado por maior seletividade e restrições fiscais do Fies, conforme discutido por Pontuschka (2016), Souza (2021) e Becker e Mendonça (2021), com potenciais reflexos sobre a organização operacional e a eficiência das instituições privadas.

Em síntese, a análise inicial do Gráfico 9 indica que as variações do Índice de Produtividade de Malmquist refletem uma combinação de fatores internos às instituições e condicionantes externos de natureza econômica e institucional. Os resultados corroboram parcialmente a literatura que aponta maior dinamismo — e maior volatilidade — das IES privadas, ao passo que relativizam interpretações simplificadoras segundo as quais as IES públicas apresentariam trajetória monotônica de ganhos de produtividade. A decomposição por etapas mostra-se, assim, fundamental para evidenciar que os avanços e limitações da produtividade não se distribuem de forma uniforme ao longo do processo produtivo educacional.

Por fim, a análise conjunta do índice de Malmquist e de seus componentes sugere que as variações observadas na eficiência ao longo do período analisado decorrem predominantemente de ajustes temporários de eficiência técnica e de escala, sem evidência de deslocamentos estruturais persistentes da fronteira

tecnológica. Articulados aos resultados das análises de eficiência estática, das variáveis ambientais e dos efeitos do financiamento estudantil, esses achados oferecem uma visão integrada do desempenho institucional no ensino superior brasileiro. No capítulo seguinte, apresentam-se as Considerações Finais, nas quais se sintetizam os principais resultados do estudo, discutem-se suas implicações para a gestão e para a formulação de políticas públicas, bem como se apontam as limitações da pesquisa e sugestões para investigações futuras.

5 CONSIDERAÇÕES FINAIS

O presente estudo teve como objetivo avaliar a eficiência técnica aplicada ao processo operacional das Instituições de Ensino Superior (IES) brasileiras, públicas federais e privadas, no período de 2019 a 2023, considerando indicadores de acesso, permanência, conclusão e qualidade educacional. Para isso, foram analisadas as diferenças de desempenho institucional entre os dois segmentos de ensino e investigada a influência de fatores contextuais e de políticas públicas de financiamento, com destaque para o papel do Fies.

Os resultados evidenciaram diferenças relevantes no padrão de eficiência entre as instituições públicas e privadas. De modo geral, as IES públicas apresentaram maior estabilidade nos níveis de eficiência ao longo do período analisado, sugerindo maior consolidação institucional e menor dependência de oscilações nas políticas de financiamento. Em contraste, as IES privadas demonstraram maior heterogeneidade de desempenho, indicando que fatores institucionais e contextuais exercem papel significativo na determinação de seus resultados operacionais. Esse achado sugere que a expansão do ensino superior privado, embora relevante para ampliar o acesso, não garante automaticamente maior eficiência educacional.

A análise do processo operacional das instituições também revelou a existência de gargalos relevantes no sistema de ensino superior, particularmente relacionados às etapas de permanência e conclusão dos estudantes. Em diversas instituições privadas, observaram-se taxas relativamente elevadas de evasão, indicando que a ampliação do acesso ao ensino superior não se traduz necessariamente em resultados educacionais eficientes. Esse resultado reforça a necessidade de políticas institucionais voltadas à permanência estudantil e ao fortalecimento da gestão acadêmica.

Outro resultado relevante refere-se à influência de fatores contextuais sobre o desempenho das instituições. Variáveis demográficas, institucionais e territoriais demonstraram influência significativa sobre os níveis de eficiência das IES, evidenciando que o ambiente socioeconômico no qual as instituições estão inseridas afeta sua capacidade de converter recursos em resultados educacionais. Assim, avaliações de desempenho institucional devem considerar não apenas aspectos internos de gestão, mas também os condicionantes estruturais e regionais que moldam o funcionamento das instituições.

No que se refere às políticas públicas de financiamento educacional, os resultados indicaram que a participação no Fies está associada a padrões distintos de eficiência entre as IES privadas. O programa tende a proporcionar maior previsibilidade de receitas e estabilidade financeira às instituições participantes, o que pode favorecer investimentos em infraestrutura e qualificação docente. Contudo, observou-se que esse efeito não é uniforme ao longo do tempo, sendo influenciado por mudanças nas regras e na disponibilidade de financiamento do programa.

Por fim, a análise da evolução da produtividade das instituições indicou que as mudanças observadas no desempenho das IES brasileiras decorrem tanto de melhorias gerenciais quanto de transformações tecnológicas e organizacionais. As instituições públicas apresentaram trajetória mais estável de evolução da produtividade, enquanto as privadas demonstraram maior volatilidade, especialmente em contextos de mudanças nas políticas de financiamento estudantil.

Do ponto de vista acadêmico, esta pesquisa contribui para a literatura sobre eficiência no ensino superior ao integrar diferentes abordagens analíticas para examinar o desempenho institucional das IES brasileiras. Ao analisar simultaneamente eficiência técnica, produtividade e fatores contextuais, o estudo amplia a compreensão sobre os determinantes do desempenho institucional no sistema educacional brasileiro.

Sob a perspectiva das políticas públicas, os resultados indicam que a eficiência educacional não depende exclusivamente do volume de recursos investidos, mas também da forma como esses recursos são alocados e geridos. Instrumentos de financiamento estudantil, como o Fies, podem desempenhar papel relevante na ampliação do acesso ao ensino superior, mas sua efetividade depende da estabilidade regulatória, da sustentabilidade fiscal e da capacidade institucional das IES de converter financiamento em melhorias efetivas de desempenho educacional.

As evidências apresentadas nesta pesquisa podem subsidiar gestores públicos e formuladores de políticas educacionais na avaliação de estratégias de financiamento do ensino superior. Em particular, os resultados oferecem elementos empíricos que podem contribuir para o aprimoramento de políticas de financiamento estudantil e para o desenvolvimento de mecanismos de monitoramento e avaliação de desempenho das instituições de ensino superior.

Como principais limitações, destaca-se a utilização de dados secundários, o que restringe o controle sobre a qualidade e a periodicidade das informações

disponíveis. Além disso, a análise concentrou-se no período de 2019 a 2023, o que limita a possibilidade de inferências sobre tendências de longo prazo no sistema de ensino superior brasileiro.

Outro aspecto relevante refere-se à ausência de indicadores qualitativos relacionados à percepção de estudantes, docentes e gestores sobre o desempenho institucional, os quais poderiam complementar a análise quantitativa da eficiência educacional.

Diante dessas limitações, pesquisas futuras podem ampliar o horizonte temporal de análise, incorporando séries históricas mais extensas que permitam avaliar a evolução de longo prazo da eficiência institucional. Também se recomenda a inclusão de indicadores qualitativos relacionados à experiência acadêmica e à qualidade percebida do ensino.

Além disso, estudos comparativos internacionais podem contribuir para compreender como diferentes modelos de financiamento e governança universitária influenciam o desempenho das instituições. Por fim, abordagens analíticas capazes de capturar de forma mais detalhada a dinâmica dos subprocessos educacionais — como ensino, pesquisa e extensão — podem ampliar a compreensão sobre os determinantes da eficiência no ensino superior.

REFERÊNCIAS

- ABRAMO, G.; D'ANGELO, C. A.; ROSATI, F. The determinants of academic career advancement: evidence from Italy. **Science and Public Policy**, [s. l.], v. 42, n. 6, p. 761-774, 2015. DOI 10.1093/scipol/scu086. Disponível em: <https://academic.oup.com/spp/article-abstract/42/6/761/2364597>. Acesso em: 27 abr. 2025.
- AFONSO, A.; SCHUKNECHT, L.; TANZI, V. Public sector efficiency: an international comparison. **Public Choice**, v. 123, n. 3/4, p. 321-347, 2005. DOI 10.1007/s11127-005-7165-2. Disponível em: <https://link.springer.com/article/10.1007/s11127-005-7165-2>. Acesso em: 12 fev. 2025.
- AGASISTI, T.; EGOROV, A.; ZINCHENKO, D.; LESHUKOV, O. Efficiency of regional higher education systems and regional economic short-run growth: empirical evidence from Russia. **Industry and Innovation**, v. 28, n. 4, p. 507-534, 2021. DOI 10.1080/13662716.2020.1738914. Disponível em: <https://www.tandfonline.com/doi/full/10.1080/13662716.2020.1738914>. Acesso em: 19 jun. 2025.
- AGASISTI, T.; PÉREZ-ESPARRELLS, C. Comparing efficiency in a cross-country perspective: the case of Italian and Spanish state universities. **Higher Education**, v. 59, n. 1, p. 85-103, 2010. DOI 10.1007/s10734-009-9235-8. Disponível em: <https://link.springer.com/article/10.1007/s10734-009-9235-8>. Acesso em: 3 set. 2025.
- ALMEIDA, A. T. C. de; RAMALHO, H. M. de B.; SOBRAL, E. F. M.; SILVA, A. F. da. Análise da eficiência dos departamentos de economia de instituições de ensino superior do Brasil. **Economia Aplicada**, v. 22, n. 1, p. 109-140, 2018. DOI 10.11606/1980-5330/ea122240. Disponível em: https://revistas.usp.br/ecoa/pt_BR/article/view/122240. Acesso em: 5 mar. 2025.
- ANDERSEN, P.; PETERSEN, N. C. A procedure for ranking efficient units in data envelopment analysis. **Management Science**, v. 39, n. 10, p. 1261-1264, 1993. DOI 10.1287/mnsc.39.10.1261. Disponível em: <https://pubsonline.informs.org/doi/abs/10.1287/mnsc.39.10.1261>. Acesso em: 17 ago. 2025.
- ANDERSSON, C.; QUIGLEY, J.; WILHELMSSON, M.; WEINER, J. Technical efficiency and productivity for higher education institutions in Sweden. **Scandinavian Journal of Educational Research**, v. 61, n. 2, p. 205-223, 2017. DOI 10.1080/00313831.2015.1120230. Disponível em: <https://www.tandfonline.com/doi/abs/10.1080/00313831.2015.1120230>. Acesso em: 29 maio 2025.
- ARBULA BLECICH, A. Efficiency evaluation of the higher education sector in Europe – Window DEA based approach. **Zbornik Veleučilišta u Rijeci**, v. 12, n. 1, p. 15-30, 2024. DOI 10.31784/zvr.12.1.9. Disponível em: <https://hrcak.srce.hr/317615>. Acesso em: 6 jan. 2025.

BANCO MUNDIAL. **Um ajuste justo: análise da eficiência e equidade do gasto público no Brasil**. Washington, DC: Banco Mundial, 2017. Disponível em: <https://www.worldbank.org/pt/country/brazil/publication/brazil-expenditure-review-report>. Acesso em: 23 jul. 2025.

BANKER, R.D.; CHANG, H. The super-efficiency procedure for outlier identification, not for ranking efficient units. **European Journal of Operational Research**, v. 175, n. 2, p. 1311-1320, 2006. DOI 10.1016/j.ejor.2005.06.028. Disponível em: <https://www.sciencedirect.com/science/article/pii/S037722170500531X>. Acesso em: 12 jan. 2026.

BANKER, R. D.; CHARNES, A.; COOPER, W. W. Some models for estimating technical and scale efficiencies in data envelopment analysis. **Management Science**, v. 30, n. 9, p. 1078-1092, 1984. DOI 10.1287/mnsc.30.9.1078. Disponível em: <https://pubsonline.informs.org/doi/abs/10.1287/mnsc.30.9.1078>. Acesso em: 18 dez. 2024.

BARRA, C.; ZOTTI, R. Measuring efficiency in higher education: an empirical study using a bootstrapped data envelopment analysis. **International Advances in Economic Research**, v. 22, n. 1, p. 11-33, 2016. DOI 10.1007/s11294-015-9558-4. Disponível em: <https://link.springer.com/article/10.1007/s11294-015-9558-4>. Acesso em: 9 out. 2025.

BEAN, J. P. Dropouts and turnover: the synthesis and test of a causal model of student attrition. **Research in Higher Education**, v. 12, n. 2, p. 155-187, 1980. DOI 10.1107/BF00976194. DOI 10.1007/BF00976194. Disponível em: <https://link.springer.com/article/10.1007/bf00976194>. Acesso em: 2 out. 2025.

BECKER, K. L.; MENDONÇA, M. J. C. Políticas de financiamento estudantil: análise de impacto do Fies no tempo de conclusão do ensino superior. **Economia e Sociedade**, v. 30, n. 2, p. 551-581, 2021. DOI 10.1590/1982-3533.2021v30n2art11. Disponível em: <https://www.scielo.br/j/ecos/a/5NKNCs9YbL48y3gM5SSM7bt/?format=html&lang=pt>. Acesso em: 14 jan. 2025.

BEHN, R. D. Why measure performance? Different purposes require different measures. **Public Administration Review**, v. 63, n. 5, p. 586-606, 2003. DOI 10.1111/1540-6210.00322. Disponível em: <https://onlinelibrary.wiley.com/doi/abs/10.1111/1540-6210.00322>. Acesso em: 05 mar. 2026.

BELLONI, J. A. **Uma metodologia de avaliação da eficiência produtiva de universidades federais brasileiras**. Tese (Doutorado) — Universidade Federal de Santa Catarina. Florianópolis. 2000. Disponível em: <http://teses.eps.ufsc.br/defesa/pdf/1757.pdf>. Acesso em: 18 jun. 2025.

BIELSCHOWSKY, C. E.; AMARAL, N. C. O custo do aluno das 2.537 instituições de educação superior brasileiras: cai um mito? **Educação & Sociedade**, v. 43, e243866, 2022. DOI 10.1590/ES.243866. Disponível em:

<https://www.scielo.br/j/es/a/3SxH5sZPhCxRKc5KBqyNZvd/?lang=pt>. Acesso em: 7 abr. 2025.

BITTENCOURT, H. R.; CASARTELLI, A. de O.; RODRIGUES, A. C. de M. Sobre o Índice Geral de Cursos (IGC). **Avaliação: Revista da Avaliação da Educação Superior (Campinas)**, v. 14, n. 3, p. 667-682, 2009. DOI 10.1590/S1414-40772009000300008. Disponível em: https://educa.fcc.org.br/scielo.php?pid=S1414-40772009000300008&script=sci_abstract. Acesso em: 30 jan. 2025.

BOGETOFT, P.; OTTO, L. **Benchmarking with dea, sfa, and r**. Springer Science & Business Media, 2010. Disponível em: <http://ndl.ethernet.edu.et/bitstream/123456789/21766/1/318.pdf>. Acesso em: 16 set. 2025.

BOUERI, R.; ROCHA, F.; RODOPOULOS, F. (org.). **Avaliação da qualidade do gasto público e mensuração da eficiência**. Brasília, DF: Secretaria do Tesouro Nacional, 2015. 463 p. Disponível em: https://www.researchgate.net/profile/Rafael-Terra/publication/315644428_Conceitos_sobre_eficiencia/links/58d7419caca2727e5eea9cb9/Conceitos-sobre-eficiencia.pdf. Acesso em: 11 mar. 2025.

BRASIL. Constituição (1988). Emenda Constitucional n. 95, de 15 de dezembro de 2016. Altera o Ato das Disposições Constitucionais Transitórias, para instituir o Novo Regime Fiscal, e dá outras providências. Brasília, DF: Presidência da República, 2016. Disponível em: http://www.planalto.gov.br/ccivil_03/constituicao/emendas/emc/emc95.htm. Acesso em: 5 mar. 2025.

BRASIL. Controladoria-Geral da União. **Portal da Transparência da União: Orçamento da Despesa Pública**. Brasília, DF: CGU, 2023a. Disponível em: <https://portaldatransparencia.gov.br/orcamento/despesas?paginacaoSimples=true&amanhoPagina=&offset=&direcaoOrdenacao=asc&orgaos=OS26000&funcaoSubfuncao=FN12%2CSB364&colunasSelecionadas=ano%2CorgaoSuperior%2CorgaoVinculado%2CunidadeOrcamentaria%2Cfuncao%2CsubFuncao%2Cprograma%2Cacao%2CcategoriaEconomica%2CgrupoDespesa%2CelementoDespesa%2Corcamentonicial%2CorcamentoAtualizado%2CorcamentoRealizado%2CvalorEmpenhado%2CpercentualRealizado>. Acesso em: 24 maio 2025.

BRASIL. Controladoria-Geral da União. **Portal da Transparência da União: Concessão de Financiamento Estudantil - FIES (Lei n. 10.260, de 2001)**. Brasília, DF: CGU, 2023b. Disponível em: <https://portaldatransparencia.gov.br/programas-e-aco-es/acao/00IG-concessao-de-financiamento-estudantil---Fies--lei-n-----de-----?ano=2021>. Acesso em: 24 maio 2025.

BRASIL. Fundo Nacional de Desenvolvimento da Educação (FNDE). **Mantenedora e IES**. Brasília, DF: Ministério da Educação, [s.d.]a. Disponível em: https://www.gov.br/fnde/pt-br/aceso-a-informacao/acoes-e-programas/financiamento/Fies/mantenedora_ies. Acesso em: 2 fev. 2025.

BRASIL. Ministério da Educação. Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep). **Índice Geral de Cursos (IGC)**. Brasília, DF:

Inep, [s.d.]b. Disponível em: <https://www.gov.br/inep/pt-br/assuntos/indicadores-educacionais/indice-geral-de-cursos-igc>. Acesso em: 21 abr. 2025.

BRASIL. Ministério da Educação. Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep). **Painel estatístico do Censo da Educação Superior**. Brasília, DF: Inep, [s.d.]c. Disponível em: <https://www.gov.br/inep/pt-br/assuntos/pesquisas-estatisticas-e-indicadores/censo-da-educacao-superior/painel-de-indicadores>. Acesso em: 21 abr. 2025.

BRASIL. Instituto Brasileiro de Geografia e Estatística (IBGE). **Estatísticas: downloads de estatísticas**. Rio de Janeiro, RJ: IBGE, [s.d.]d. Disponível em: <https://www.ibge.gov.br/estatisticas/downloads-estatisticas.html>. Acesso em: 21 abr. 2025.

BRASIL. Instituto Brasileiro de Geografia e Estatística (IBGE). **Produto Interno Bruto dos municípios**. Rio de Janeiro, RJ: IBGE, [s.d.]e. Disponível em: <https://www.ibge.gov.br/estatisticas/economicas/contas-nacionais/9088-produto-interno-bruto-dos-municipios.html>. Acesso em: 24 abr. 2025.

BREIMAN, L. Random forests. **Machine Learning**, v. 45, n. 1, p. 5-32, 2001. DOI 10.1023/A:1010933404324. Disponível em: <https://link.springer.com/article/10.1023/a:1010933404324>. Acesso em: 28 fev. 2025.

BRZEZICKI, L. The efficiency of public and private higher education institutions in Poland. **Gospodarka Narodowa. The Polish Journal of Economics**, v. 304, n. 4, p. 33-51, 2020. DOI 10.33119/GN/128218. Disponível em: <https://www.ceeol.com/search/article-detail?id=918979>. Acesso em: 6 jun. 2025.

CAO, S.; QIU, Y.; UNARTA, I. C.; GOONETILLEKE, E. C.; HUANG, X. The ion-dipole correction of the 3DRISM solvation model to accurately compute water distributions around negatively charged biomolecules. **The Journal of Physical Chemistry B**, v. 126, n. 43, p. 8632-8645, 2022. DOI 10.1021/ac.jp cb.2c04431. Disponível em: <https://pubs.acs.org/doi/abs/10.1021/acs.jp cb.2c04431>. Acesso em: 25 mar. 2025.

CARDOSO, R. C. L.; SAMPAIO, H. Estudantes universitários e o trabalho. **Revista Brasileira de Ciências Sociais**, v. 26, n. 9, p. 30-50, 1994. Disponível em: https://www.kufunda.net/publicdocs/rbcs26_03.pdf. Acesso em: 10 jul. 2025.

CAVES, D. W.; CHRISTENSEN, L. R.; DIEWERT, W. E. The economic theory of index numbers and the measurement of input, output, and productivity. **Econometrica**, v. 50, n. 6, p. 1393-1414, 1982. DOI 10.2307/1913388. Disponível em: <https://www.jstor.org/stable/1913388>. Acesso em: 5 out. 2025.

CHARNES, A.; COOPER, W. W.; RHODES, E. Measuring the efficiency of decision-making units. **European Journal of Operational Research**, v. 2, n. 6, p. 429-444, 1978. DOI 10.1016/0377-2217(78)90138-8. Disponível em: <https://www.sciencedirect.com/science/article/pii/0377221778901388>. Acesso em: 22 out. 2024.

CHEN, T.; GUESTIN, C. XGBoost: a scalable tree boosting system. In: **Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16)**, 2016, San Francisco, CA. New York: ACM, 2016. p. 785-794. DOI 10.1145/2939672.2939785. Disponível em: <https://dl.acm.org/doi/10.1145/2939672.2939785>. Acesso em: 18 jan. 2025.

CHEN, J.; ZHOU, X.; YAO, J.; TANG, S.-K. Application of machine learning in higher education to predict students' performance, learning engagement and self-efficacy: a systematic literature review. **Asian Education and Development Studies**, v. 14, n. 2, p. 205-240, 2025. DOI 10.1108/AEDS-08-2024-0166. Disponível em: <https://www.emerald.com/aeds/article-abstract/14/2/205/1240318/Application-of-machine-learning-in-higher?redirectedFrom=fulltext>. Acesso em: 2 out. 2025.

CIVERA, A.; LEHMANN, E. E.; PALEARI, S.; STOCKINGER, S. A. E. Higher education policy: why hope for quality when rewarding quantity? **Research Policy**, v. 49, n. 8, p. 104083, 2020. DOI 10.1016/j.respol.2020.104083. Disponível em: <https://www.sciencedirect.com/science/article/pii/S004873332030161X>. Acesso em: 11 nov. 2024.

COELLI, T.; RAO, D. S. P.; O'DONNELL, C. J.; BATTESE, G. E. **An introduction to efficiency and productivity analysis**. Boston, MA: Springer US, 2005. Disponível em: <http://download.e-bookshelf.de/download/0000/0005/80/L-G-0000000580-0002368702.pdf>. Acesso em: 26 jul. 2025.

CONOVER, W. J. **Practical nonparametric statistics**. 3. ed. New York: John Wiley & Sons, 1999. Disponível em: https://books.google.com.br/books?hl=pt-BR&lr=&id=n_39DwAAQBAJ&oi=fnd&pg=PA1&dq=Practical+nonparametric+statistics&ots=LagtV0knH6&sig=OMctRBmdxj8Ko54cZW3S1ZJEkYw&redir_esc=y#v=onepage&q=Practical%20nonparametric%20statistics&f=false. Acesso em: 15 abr. 2025.

CONTRERAS, I.; LOZANO, S. Allocating additional resources to public universities: a DEA bargaining approach. **Socio-Economic Planning Sciences**, v. 71, p. 100752, 2020. DOI 10.1016/j.seps.2019.100752. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0038012119302009>. Acesso em: 13 mar. 2025.

COOPER, W. W.; SEIFORD, L. M.; TONE, K. **Data envelopment analysis: a comprehensive text with models, applications, references and DEA-Solver software**. New York: Springer, 2007. Disponível em: <https://link.springer.com/book/10.1007/b109347>. Acesso em: 28 ago. 2025.

COOPER, W. W.; SEIFORD, L. M.; ZHU, J. **Handbook on data envelopment analysis**. 2. ed. New York: Springer, 2011. Disponível em: https://books.google.com.br/books?hl=pt-BR&lr=&id=fBbrAjpSpLYC&oi=fnd&pg=PR3&dq=Handbook+on+data+envelopment+analysis&ots=vdlw5rGA-U&sig=vRz-G1VAj0dXSBxsHZWWk3NHUjk&redir_esc=y#v=onepage&q=Handbook%20on%20data%20envelopment%20analysis&f=false. Acesso em: 4 maio 2025.

COSTA, E. M.; SOUZA, H. R.; RAMOS, F. S.; SILVA, J. L. M. da. Eficiência e desempenho no ensino superior: uma análise da fronteira de produção educacional das IFES brasileiras. **Revista de Economia Contemporânea**, v. 16, n. 3, p. 415-440, 2012. DOI 10.1590/S1415-98482012000300003. Disponível em: <https://www.scielo.br/j/rec/a/VMp6StXRyHq9f7MYgS53BPb/?format=html&lang=pt>. Acesso em: 2 dez. 2024.

CUNHA, L. A. **A universidade temporã: o ensino superior, da Colônia à Era Vargas**. 3. ed. São Paulo: Unesp, 2007. DOI 10.7476/9788539304578. Disponível em: https://books.google.com.br/books?hl=pt-BR&lr=&id=5C6_laEWdUoC&oi=fnd&pg=PA9&dq=A+universidade+tempor%C3%A3:+o+ensino+superior,+da+Col%C3%B4nia+%C3%A0+Era+Vargas&ots=oLa2PqvGiN&sig=3DtArl0Ch9hh3rqCfuE9wT5jg1w&redir_esc=y#v=onepage&q=A%20universidade%20tempor%C3%A3%3A%20o%20ensino%20superior%2C%20da%20Col%C3%B4nia%20%C3%A0%20Era%20Vargas&f=false. Acesso em: 15 mai. 2025.

DALLABRIDA, N. A reforma Francisco Campos e a modernização nacionalizada do ensino secundário. **Educação**, v. 32, n. 2, p. 185-191, 2009. DOI 10.5902/19846444. Disponível em: https://educa.fcc.org.br/scielo.php?pid=S1981-25822009000200011&script=sci_abstract. Acesso em: 4 abr. 2025.

EMROUZNEJAD, A.; YANG, G. L. A survey and analysis of the first 40 years of scholarly literature in DEA: 1978–2016. **Socio-Economic Planning Sciences**, v. 61, p. 4-8, 2018. DOI 10.1016/j.seps.2017.01.008. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0038012117300174>. Acesso em: 27 mar. 2025.

ENTANI, T.; MAEDA, Y.; TANAKA, H. Dual models of interval DEA and its extensions to interval data. **European Journal of Operational Research**, v. 136, n. 1, p. 32-45, 2002. DOI 10.1016/S0377-2217(01)00055-8. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0377221701000558>. Acesso em: 9 nov. 2024.

FAÇANHA, L. O.; MARINHO, A. Instituições Federais de Ensino Superior: modelos de financiamento e o incentivo à eficiência. **Revista Brasileira de Economia**, v. 53, n. 3, p. 357-386, 1999. Disponível em: <https://periodicos.fgv.br/rbe/article/download/756/8114>. Acesso em: 28 nov. 2024.

FAÇANHA, L. O.; MARINHO, A. Instituições de ensino superior governamentais e particulares: avaliação comparativa da eficiência. **Revista de Administração Pública**, v. 35, n. 6, p. 83-105, 2001. Disponível em: <https://periodicos.fgv.br/rap/article/view/6416>. Acesso em: 9 jan. 2025.

FALQUETTO, A. M.; TAKASAGO, M.; PEÑA, C. R.; ARAÚJO NETO, L. M. de; SALES, I. C. H. Avaliação da eficiência dos programas de economia no país contemplados com o Proex e o Proap. **Race: Revista de Administração, Contabilidade e Economia**, v. 17, n. 1, p. 333-364, 2018. DOI 10.18593/race.v17i1.16299. Disponível em: <https://dialnet.unirioja.es/servlet/articulo?codigo=6558340>. Acesso em: 18 mar. 2025.

FÄRE, R.; GROSSKOPF, S. Network DEA. **Socio-Economic Planning Sciences**, v. 34, n. 1, p. 35-49, 2000. DOI 10.1016/S0038-0121(99)00012-9. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0038012199000129>. Acesso em: 8 ago. 2025.

FÄRE, R.; GROSSKOPF, S.; NORRIS, M.; ZHANG, Z. Productivity growth, technical progress, and efficiency change in industrialized countries. **The American Economic Review**, v. 84, n. 1, p. 66-83, 1994. Disponível em: <http://www.jstor.org/stable/2117971>. Acesso em: 23 set. 2025.

FARRELL, M. J. The measurement of productive efficiency. **Journal of the royal statistical society series a: statistics in society**, v. 120, n. 3, p. 253-290, 1957. DOI 10.2307/2343100. Disponível em: <https://academic.oup.com/jrssa/article-abstract/120/3/253/7101561>. Acesso em: 7 ago. 2025.

FÁVERO, M. de L. de A. A universidade no Brasil: das origens à Reforma Universitária de 1968. **Educar em Revista**, n. 28, p. 17-36, 2006. DOI 10.1590/S0104-40602006000100003. Disponível em: <https://www.scielo.br/j/er/a/vBpNhyd7mBY3HjK46nHVvF/?lang=pt>. Acesso em: 12 jun. 2025.

FELLER, W. On the Kolmogorov–Smirnov limit theorems for empirical distributions. **The Annals of Mathematical Statistics**, v. 19, n. 2, p. 177-189, 1948. DOI 10.1214/aoms/1177730243. Disponível em: <https://projecteuclid.org/journals/annals-of-mathematical-statistics/volume-19/issue-2/On-the-Kolmogorov-Smirnov-Limit-Theorems-for-Empirical-Distributions/10.1214/aoms/1177730243.full>. Acesso em: 2 out. 2025.

FERRARI, S.; CRIBARI-NETO, F. Beta regression for modeling rates and proportions. **Journal of Applied Statistics**, v. 31, n. 7, p. 799-815, 2004. DOI 10.1080/0266476042000214501. Disponível em: <https://www.tandfonline.com/doi/abs/10.1080/0266476042000214501>. Acesso em: 11 ago. 2025.

FRIED, H. O.; LOVELL, C. A. K.; SCHMIDT, S. S. **The measurement of productive efficiency and productivity growth**. Oxford University Press, 2008. Disponível em: <https://books.google.tt/books?id=PBITDAAAQBAJ&printsec=frontcover&hl=pt-BR>. Acesso em: 30 set. 2025. *E-book*.

GARCIA, C. G.; CATANI, A. M. La reforma de la educación superior en Venezuela desde una perspectiva comparada. In: CATANI, A. M. (Org.). **Congresso Internacional de Políticas de Educação Superior na América Latina no Limiar do Século XXI**, Recife. 1997. p. 189-201. Disponível em: https://home.ufam.edu.br/andersonlfc/Cole%C3%A7%C3%A3o%20-%20Os%20Economistas/Afr%25C3%25A2nio_Mendes_Catani_-_Am%25C3%25A9rica_Latina_Impasses_E_Alternativas.pdf. Acesso em: 18 jul. 2025.

GIMÉNEZ, V.; PRIOR, D.; THIEME, C. Technical efficiency, managerial efficiency and objective-setting in the educational system: an international comparison. **Journal**

of the **Operational Research Society**, v. 58, n. 8, p. 996-1007, 2007. DOI 10.1057/palgrave.jors.2602213. Disponível em: <https://www.tandfonline.com/doi/abs/10.1057/palgrave.jors.2602213>. Acesso em: 15 jul. 2025.

GOMES JÚNIOR, S.; MORENO, P.; MELLO, J. C. C. S.; ANGULO-MEZA, L. Utilização de modelo Network DEA na avaliação de cursos de pós-graduação stricto sensu em engenharia. **Proceedings of the XVII Simpósio de Pesquisa Operacional e Logística da Marinha—SPOLM, Rio de Janeiro, Brasil**, p. 6-9, 2014. DOI 10.13140/2.1.4308.2884. Disponível em: <https://pdf.blucher.com.br/marineengineeringproceedings/spolm2014/126105.pdf>. Acesso em: 18 jul. 2025.

GUCCIO, C.; MARTORANA, M. F.; MONACO, L. Evaluating the impact of the Bologna Process on the efficiency convergence of Italian universities: a non-parametric frontier approach. **Journal of Productivity Analysis**, v. 45, n. 3, p. 275-298, 2016. DOI 10.1007/s11123-015-0459-6. Disponível em: <https://link.springer.com/article/10.1007/s11123-015-0459-6>. Acesso em: 28 ago. 2025.

GUIRONNET, J. P.; PEYPOCH, N. The geographical efficiency of education and research: the ranking of US universities. **Socio-Economic Planning Sciences**, v. 62, p. 44-55, 2018. DOI 10.1016/j.seps.2017.07.003. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0038012117300241>. Acesso em: 21 set. 2025.

GUL, M.; YUCESAN, M. Performance evaluation of Turkish universities by an integrated Bayesian BWM-TOPSIS model. **Socio-Economic Planning Sciences**, v. 80, art. 101173, 2022. DOI 10.1016/j.seps.2021.101173. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0038012121001658>. Acesso em: 26 ago. 2025.

HOFF, A. Second stage DEA: comparison of approaches for modelling the DEA score. **European Journal of Operational Research**, v. 181, n. 1, p. 425-435, 2007. DOI 10.1016/j.ejor.2006.05.019. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0377221706003821>. Acesso em: 7 set. 2025.

JANIESCH, C.; ZSCHECH, P.; HEINRICH, K. Machine learning and deep learning. **Electronic Markets**, v. 31, p. 685-695, 2021. DOI 10.1007/s12525-021-00475-2. Disponível em: <https://link.springer.com/article/10.1007/s12525-021-00475-2>. Acesso em: 18 set. 2025.

JOHNES, G.; RUGGIERO, J. Revenue efficiency in higher education institutions under imperfect competition. **Public Policy and Administration**, v. 32, n. 4, p. 282-295, 2017. DOI 10.1177/0952076716652935. Disponível em: <https://journals.sagepub.com/doi/abs/10.1177/0952076716652935>. Acesso em: 9 set. 2025.

JOHNES, J. Measuring efficiency: a comparison of multilevel modelling and data envelopment analysis in the context of higher education. **Bulletin of Economic Research**, v. 58, n. 2, p. 75-104, 2006. DOI 10.1111/j.0307-3378.2006.00238.x. Disponível em: <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.0307-3378.2006.00238.x>. Acesso em: 30 ago. 2025.

KAO, C. Network data envelopment analysis: a review. **European Journal of Operational Research**, v. 239, n. 1, p. 1-16, 2014. DOI 10.1016/j.ejor.2014.02.039. Disponível em: <https://www.sciencedirect.com/science/article/pii/S037722171400174X>. Acesso em: 25 ago. 2025.

KOOPMANS, T. C. **Activity analysis of production and allocation**. New York: John Wiley & Sons, 1951. (Cowles Commission Monograph, n. 13). Disponível em: <https://cowles.yale.edu/sites/default/files/2022-09/m13-all.pdf>. Acesso em: 30 ago. 2025.

KREJNUŠ, M.; ŠTOFKOVÁ, J.; ŠTOFKOVÁ, K. R.; BINASOVÁ, V. The use of the DEA method for measuring the efficiency of electronic public administration as part of the digitization of the economy and society. **Applied Sciences**, v. 13, n. 6, art. 3672, 2023. DOI 10.3390/app13063672. Disponível em: <https://www.mdpi.com/2076-3417/13/6/3672>. Acesso em: 12 ago. 2025.

KUHN, M.; JOHNSON, K. **Applied predictive modeling**. New York: Springer, 2013. Disponível em: https://www.ic.unicamp.br/~wainer/cursos/1s2021/432/2013_Book_AppliedPredictiveModeling.pdf. Acesso em: 6 set. 2025.

LEE, C. **Malmquist productivity analysis using DEA frontier in Stata**. In: STATA CONFERENCE, 2011, Chicago. Apresentação Chicago: StataCorp, 2011. Disponível em: http://fmwww.bc.edu/repec/chic2011/chi11_lee.ppt. Acesso em: 28 set. 2025.

LEHMANN, E. E.; MEOLI, M.; PALEARI, S.; STOCKINGER, A. A. E. Approaching effects of the economic crisis on university efficiency: a comparative study of Germany and Italy. **Eurasian Business Review**, v. 8, n. 1, p. 37-54, 2018. DOI 10.1007/s40821-017-0091-7. Disponível em: <https://link.springer.com/article/10.1007/s40821-017-0091-7>. Acesso em: 22 ago. 2025.

LEWIS, H. F.; SEXTON, T. R. Network DEA: efficiency analysis of organizations with complex internal structure. **Computers & Operations Research**, v. 31, n. 9, p. 1365-1410, 2004. DOI 10.1016/S0305-0548(03)00095-9. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0305054803000959>. Acesso em: 5 set. 2025.

LIU, J. S.; LU, L. Y. Y.; LU, W. M.; LIN, B. J. Y. A survey of DEA applications. **Omega**, v. 41, n. 5, p. 893-902, 2013. DOI 10.1016/j.omega.2012.11.004. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0305048312002186>. Acesso em: 9 set. 2025.

LOVELL, C. A. K. Production frontiers and productive efficiency. In: FRIED, H. O.; LOVELL, C. A. K.; SCHMIDT, S. S. (eds.). **The measurement of productive efficiency: techniques and applications**. New York: Oxford University Press, 1993. p. 3-67. Disponível em:

https://www.researchgate.net/publication/239060004_Production_Frontiers_and_Productive_Efficiency. Acesso em: 12 ago. 2025.

MACEDO, M. A. S.; ROSADAS, L. A. S. Modelagem DEA na formação do preço de venda. **Revista de Administração FACES Journal**, v. 5, n. 2, p. 11-24, 2006. DOI 10.21714/1984-6975FACES2006V5N2ART70. Disponível em:

<https://revista.fumec.br/index.php/facesp/article/view/70>. Acesso em: 3 set. 2025.

MALMQUIST, S. Index numbers and indifference surfaces. **Trabajos de Estadística**, v. 4, n. 2, p. 209-242, 1953. DOI 10.1007/BF03006863. Disponível em: <https://link.springer.com/article/10.1007/BF03006863>. Acesso em: 19 set. 2025.

MARIANO, Enzo B. Conceitos básicos de análise de eficiência produtiva. **XII Simpósio de Engenharia de Produção, SIMPEP**, 2007. DOI

10.13140/2.1.3202.1441. Disponível em:

https://www.researchgate.net/publication/257397765_Conceitos_Basicos_de_Analise_de_Eficiencia_produtiva. Acesso em: 31 ago. 2025.

MARINHO, A.; RESENDE, M.; FAÇANHA, L. Brazilian federal universities: relative efficiency evaluation and data envelopment analysis. **Revista Brasileira de Economia**, v. 51, n. 4, p. 489-508, 1997. Disponível em:

<https://periodicos.fgv.br/rbe/article/download/701/8066>. Acesso em: 7 set. 2025.

MARSAGLIA, G.; TSANG, W. W.; WANG, J. Evaluating Kolmogorov's distribution. **Journal of Statistical Software**, v. 8, n. 18, p. 1-4, 2003. DOI

10.18637/jss.v008.i18. Disponível em: <https://www.jstatsoft.org/article/view/v008i18>. Acesso em: 6 set. 2025.

MARTÍNEZ-CAMPILLO, A.; FERNÁNDEZ-SANTOS, Y. The impact of the economic crisis on the (in)efficiency of public higher education institutions in Southern Europe: the case of Spanish universities. **Socio-Economic Planning Sciences**, v. 71, p. 100771, 2020. DOI 10.1016/j.seps.2019.100771. Disponível em:

<https://www.sciencedirect.com/science/article/pii/S0038012119302538>. Acesso em: 11 set. 2025.

MARTINS, A. C. P. Ensino superior no Brasil: da descoberta aos dias atuais. **Acta Cirúrgica Brasileira**, v. 17, supl. 3, p. 4-6, 2002. DOI 10.1590/S0102-86502002000900001. Disponível em:

<https://www.scielo.br/j/acb/a/8jQH56v8cDtWGZ8yZdYjHHQ/?format=html&lang=pt>. Acesso em: 30 set. 2025.

MARTINS, C. B. A reforma universitária de 1968 e a abertura para o ensino superior privado no Brasil. **Educação & Sociedade**, v. 30, n. 106, p. 15-35, 2009. DOI 10.1590/S0101-73302009000100002. Disponível em:

<https://www.scielo.br/j/es/a/RKsKcwfYc6QVFBHy4nvJzHt/?format=html&lang=pt>. Acesso em: 20 set. 2025.

MARTINS FILHO, J. R. **Movimento estudantil e ditadura militar (1964-1968)**. Campinas: Papyrus, 1987. 213 p.

MAS-COLELL, A.; WHINSTON, M. D.; GREEN, J. R. **Microeconomic theory**. New York: Oxford University Press, 1995. Disponível em: https://moodle.umontpellier.fr/pluginfile.php/904638/mod_resource/content/1/Mas-colell%2C%20Whinston%2C%20Green%20-%201995%20-%20Oxford%20University%20Press.pdf. Acesso em: 17 set. 2025.

MCDONALD, J. Using least squares and Tobit in second stage DEA efficiency analyses. **European Journal of Operational Research**, v. 197, n. 2, p. 792-798, 2009. DOI 10.1016/j.ejor.2008.07.039. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0377221708006693>. Acesso em: 29 ago. 2025.

MEZA, A. L.; NETO, L. B.; MELLO, J. C. C. S.; GOMES, E. G. Curso de análise de envoltória de dados. **XXXVII Simpósio brasileiro de pesquisa operacional**, p. 20520-2547, 2005. Disponível em: https://www.academia.edu/26271204/CURSO_DE_AN%C3%81LISE_DE_ENVOLT%C3%93RIA_DE_DADOS. Acesso em: 22 mar. 2025.

MILAN, G. S.; SACILOTO, E. B.; LARENTIS, F.; DE TONI, D. **REAd. Revista Eletrônica de Administração (Porto Alegre)**, v. 22, n. 2, p. 419-453, 2016. DOI 10.1590/1413-2311.0982015.57273. Disponível em: <https://www.scielo.br/j/read/a/yLPZMqJwFJMR7PfljYZKk5t/?lang=pt&format=html>. Acesso em: 8 set. 2025.

MONCAYO-MARTÍNEZ, L. A.; RAMÍREZ-NAFARRATE, A.; HERNÁNDEZ-BALDERRAMA, M. G. Evaluation of public HEI on teaching, research, and knowledge dissemination by Data Envelopment Analysis. **Socio-Economic Planning Sciences**, v. 69, p. 100718, 2020. DOI 10.1016/j.seps.2019.06.003. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0038012119300205>. Acesso em: 7 out. 2025.

MOTA JÚNIOR, A. D.; TORRES, H. C. Contexto histórico da educação superior no Brasil. In: CONGRESSO NACIONAL DE EDUCAÇÃO (CONEDU), 3., 2016, João Pessoa. **Anais...** João Pessoa: Editora Realize, 2016. Disponível em: <https://www.editorarealize.com.br/artigo/visualizar/20564>. Acesso em: 12 set. 2025.

MOYNIHAN, D. P. **The dynamics of performance management: constructing information and reform**. Washington, DC: Georgetown University Press, 2008. Disponível em: https://books.google.com.br/books?hl=pt-BR&lr=&id=4-YZW-duoZ8C&oi=fnd&pg=PR3&dq=The+dynamics+of+performance+management:+constructing+information+and+reform.+Washington,+DC:+Georgetown+University+Press,+2008.&ots=CdSCyAP-u7&sig=1Npfodjw6H-5CPRwGXIn_V3qCvY&redir_esc=y#v=onepage&q=The%20dynamics%20of%20performance%20management%3A%20constructing%20information%20and%20reform.

%20Washington%2C%20DC%3A%20Georgetown%20University%20Press%2C%202008.&f=false. Acesso em: 05 mar. 2026.

NAVARRO-CHÁVEZ, J. C. L.; GÓMEZ MONGE, R.; TORRES HERNÁNDEZ, Z. Universities in Mexico: a measure of its efficiency through data envelopment analysis with bootstrap. **Acta Universitaria**, v. 26, n. 6, p. 60-69, 2017. DOI 10.15174/au.2016.911. Disponível em: https://www.scielo.org.mx/scielo.php?pid=S0188-62662016000600060&script=sci_abstract&lng=en. Acesso em: 20 set. 2025.

NAZARKO, J.; SAPARAUSKAS, J. Application of DEA method in efficiency evaluation of public higher education institutions. **Technological and Economic Development of Economy**, v. 20, n. 1, p. 25-44, 2014. DOI 10.3846/20294913.2014.837116. Disponível em: <https://www.tandfonline.com/doi/abs/10.3846/20294913.2014.837116>. Acesso em: 25 set. 2025.

NISHIMURA, J.; OKAMURO, H. Knowledge and rent spillovers through government-sponsored R&D consortia. **Science and Public Policy**, v. 43, n. 2, p. 207-225, 2016. DOI 10.1093/scipol/scv028. Disponível em: <https://academic.oup.com/spp/article-abstract/43/2/207/2414083>. Acesso em: 6 set. 2025.

OLIVEIRA, E. P.; MORAIS, S. P. As instituições de educação superior no Brasil e a influência das políticas neoliberais no corte de financiamento: experiência da Universidade Federal de Uberlândia (UFU). **Humanidades & Tecnologia (Finom)**, v. 56, n. 1, p. 59-70, 2025. DOI 10.5281/zenodo.14861966. Disponível em: <https://zenodo.org/records/14861966>. Acesso em: 5 set. 2025.

OLIVEIRA, J. P. de; BITTAR, M.; LEMOS, J. R. Políticas educacionais: cenários e projetos sociais. **Revista de Educação Pública**, v. 19, n. 40, p. 247-267, 2010. Disponível em: <https://repositorio.bc.ufg.br/bitstreams/fd1f5e6e-562e-43cf-827c-c6c61ddffedc/download>. Acesso em: 28 set. 2025.

PARTEKA, A.; WOLSZCZAK-DERLACZ, J. Dynamics of productivity in higher education: cross-European evidence based on bootstrapped Malmquist indices. **Journal of Productivity Analysis**, v. 40, n. 1, p. 67-82, 2013. DOI 10.1007/s11123-012-0320-0. Disponível em: <https://link.springer.com/article/10.1007/s11123-012-0320-0>. Acesso em: 8 nov. 2024.

PASTOR, J. T.; RUIZ, J. L.; SIRVENT, I. An enhanced DEA Russell graph efficiency measure. **European Journal of Operational Research**, v. 115, n. 3, p. 596-607, 1999. DOI 10.1016/S0377-2217(98)00098-8. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0377221798000988>. Acesso em: 14 set. 2025.

PAULA, M. F. A formação universitária no Brasil: concepções e influências. **Avaliação: Revista da Avaliação da Educação Superior (Campinas)**, v. 14, n. 1, p. 71-84, 2009. DOI 10.1590/S1414-40772009000100005. Disponível em: <https://www.scielo.br/j/aval/a/JHz4fHXBbzRXz3Xnk4VVrSw/?format=html&lang=pt>. Acesso em: 30 set. 2025.

PEREIRA, A. B. **A eficiência dos programas de pós-graduação em ciências contábeis, administração e turismo e seus determinantes**. 2023. 177 f., il. Dissertação (Mestrado em Ciências Contábeis) — Universidade de Brasília, Brasília, 2023. Disponível em: <https://repositorio.unb.br/handle/10482/50177>. Acesso em: 4 set. 2025.

PONTUSCHKA, R. **Avaliação de impacto do Fies**. 2016. Dissertação (Mestrado em Economia Aplicada) - Escola Superior de Agricultura Luiz de Queiroz, Universidade de São Paulo, Piracicaba, 2016. Disponível em: <https://doi.org/10.11606/D.11.2016.tde-03102016-163218>. Acesso em: 19 ago. 2025.

R CORE TEAM. R: a language and environment for statistical computing. **R Foundation for Statistical Computing**, 2023. Disponível em: <https://www.R-project.org/>. Acesso em: 10 set. 2025.

RAY, S. C. **Data envelopment analysis: theory and techniques for economics and operations research**. Cambridge University Press, 2004. Disponível em: <https://doi.org/10.1017/CBO9780511606731>. Acesso em: 6 set. 2025.

REIS, D. H. de O.; CARDOSO, M.; ADAMCZYK, F. **Financiamento de instituições públicas de ensino superior: análise exploratória**. Brasília: Escola Nacional de Administração Pública (ENAP), 2022. Disponível em: <https://repositorio.enap.gov.br/jspui/bitstream/1/7053/4/2022.05.03%20-%20Financiamento%20de%20institui%C3%A7%C3%B5es%20p%C3%ABlicas%20de%20ensino%20superior%20-%20an%C3%A1lise%20explorat%C3%B3ria%20-%20P2.pdf>. Acesso em: 5 mar. 2026.

ROSANO-PEÑA, C. Um modelo de avaliação da eficiência da administração pública através do método Análise Envoltória de Dados (DEA). **Revista de Administração Contemporânea**, v. 12, n. 1, p. 83-106, 2008. DOI 10.1590/S1415-65552008000100005. Disponível em: <https://www.scielo.br/j/rac/a/TdB6TqzwQyBXhB7qQgptsTf/?lang=pt>. Acesso em: 1 set. 2025.

ROSANO-PEÑA, C.; ALBUQUERQUE, P. H. M.; CARVALHO, J. M. A eficiência dos gastos públicos em educação: evidências georreferenciadas nos municípios goianos. **Economia Aplicada**, v. 16, n. 3, p. 421-443, 2012. DOI 10.1590/S1413-80502012000300004. Disponível em: <https://www.scielo.br/j/eco/a/BSbbWTG7YwF84PYW3drzLdh/?lang=pt>. Acesso em: 2 set. 2025.

ROSANO-PEÑA, C.; GOMES, E. B. P. Eficiência e produtividade no setor público: conceitos e medidas. In: MADURO-ABREU, Alexandre (Org.). **Gestão judiciária: conteúdos e disciplina**. Brasília: Editora IABS, 2018. p. 188-249. Disponível em: <https://repositorio.unb.br/handle/10482/32449>. Acesso em: 14 set. 2025.

SAMPAIO, H. **Ensino superior no Brasil: o setor privado**. Cadernos de Pesquisa, n. 111, p. 213-213, 2000. Disponível em:

https://www.revistaensinosuperior.gr.unicamp.br/edicoes/ed04_outubro2011/05_ARTIGO_PRINCIPAL.pdf. Acesso em: 18 set. 2025.

SANTANA, J. R. de; TEIXEIRA, A. L. da S.; RAPINI, M. S.; ESPERIDIÃO, F. Financiamento público à inovação no Brasil: contribuição para uma distribuição regional mais equilibrada? **Planejamento e Políticas Públicas**, v. 52, p. 356-372, 2019. Disponível em: <https://www.ipea.gov.br/ppp/index.php/PPP/article/view/796>. Acesso em: 13 ago. 2025.

SANTOS, J. P. A. dos; SILVA JÚNIOR, L. H. da; NUNES, A. Eficiência nas universidades federais brasileiras – um estudo por meio da análise envoltória de dados (DEA). **Desenvolvimento em Questão**, [S. l.], v. 20, n. 58, p. e12286, 2022. DOI 10.21527/2237-6453.2022.58.12286. Disponível em: <https://www.revistas.unijui.edu.br/index.php/desenvolvimentoemquestao/article/view/12286>. Acesso em: 9 set. 2025.

SARKER, I. H. Machine learning: algorithms, real-world applications and research directions. **SN Computer Science**, v. 2, n. 3, art. 160, 2021. Disponível em: <https://doi.org/10.1007/s42979-021-00592-x>. Acesso em: 25 set. 2025.

SAVOIE, A.; MORIN, E. L'erreur managériale: confondre les causes, les problèmes et les solutions en matière d'efficacité organisationnelle. **Gestion**, v. 26, n. 1, p. 10-11, 2001. DOI 10.3917/riges.261.0010. Disponível em: https://shs.cairn.info/article/RIGES_261_0010. Acesso em: 3 out. 2025.

SHARPE, A. Productivity concepts, trends and prospects: an overview. In: BANTING, K.; SHARPE, A.; ST-HILAIRE, F. (eds.). **The review of economic performance and social progress 2002: towards a social understanding of productivity**. Montreal: Institute for Research on Public Policy; Centre for the Study of Living Standards, 2002. p. 31-50. Disponível em: <http://www.csls.ca/repsp/2/andrewsharpe.pdf>. Acesso em: 10 set. 2025.

SHEPHARD, R. W. **Theory of cost and production functions**. Princeton: Princeton University Press, 1970. Disponível em: https://api.pageplace.de/preview/DT0400.9781400871087_A26071907/preview-9781400871087_A26071907.pdf. Acesso em: 29 set. 2025.

SILVA, C. A. da; ROSA, F. S. da. Eficiência das universidades federais brasileiras. **Avaliação: Revista da Avaliação da Educação Superior (Campinas)**, v. 27, n. 1, p. 137-158, 2022. DOI 10.1590/S1414-40772022000100008. Disponível em: <https://www.scielo.br/j/aval/a/6Mtr4dzx4PM6vy8Sy58THQH/?lang=pt>. Acesso em: 14 set. 2025.

SIMAR, L. A general methodology for bootstrapping in non-parametric frontier models. **Journal of Applied Statistics**, v. 27, n. 6, 2000. DOI 10.1080/02664760050081951. Disponível em: <https://www.tandfonline.com/doi/abs/10.1080/02664760050081951>. Acesso em: 18 set. 2025.

SIMAR, L.; WILSON, P. W. Estimation and inference in two-stage, semi-parametric models of production processes. **Journal of Econometrics**, v. 136, n. 1, p. 31-64, 2007. DOI 10.1016/j.jeconom.2005.07.009. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0304407605001594>. Acesso em: 6 out. 2025.

SIMAR, L.; WILSON, P. W. Sensitivity analysis of efficiency scores: how to bootstrap in nonparametric frontier models. **Management Science**, v. 44, n. 1, p. 49-61, 1998. DOI 10.1287/mnsc.44.1.49. Disponível em: <http://www.jstor.org/stable/2634426>. Acesso em: 22 ago. 2025.

SIMAR, L.; WILSON, P. W. Statistical approaches for non-parametric frontier models: a guided tour. **International Statistical Review**, v. 83, n. 1, p. 77-110, 2015. DOI 10.1111/insr.12056. Disponível em: <https://onlinelibrary.wiley.com/doi/abs/10.1111/insr.12056>. Acesso em: 24 ago. 2025.

SIMAR, L.; WILSON, P. W. Non-parametric tests of returns to scale. **European Journal of Operational Research**, v. 139, n. 1, p. 115-132, 2002. DOI 10.1016/S0377-2217(01)00167-9. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0377221701001679>. Acesso em: 16 set. 2025.

SIQUEIRA, J. S. **Eficiência das universidades públicas federais brasileiras**: um estudo com foco no projeto REUNI. 2015. 93 f. Dissertação (Mestrado) — Programa Multi-institucional e Inter-Regional de Pós-Graduação em Ciências Contábeis, Universidade de Brasília, Universidade Federal da Paraíba, Universidade Federal do Rio Grande do Norte, João Pessoa, 2015. Disponível em: <http://repositorio.unb.br/handle/10482/18244>. Acesso em: 26 ago. 2025.

SOUZA, M. L. S. **Programa de Financiamento Estudantil (FIES) no contexto do Plano Nacional de Educação (2014-2024)**: uma análise da expansão e privatização do ensino superior no Brasil e no Rio Grande do Norte. 2021. 188 f. Dissertação (Mestrado em Educação) – Centro de Educação, Universidade Federal do Rio Grande do Norte, Natal, 2021. Disponível em: <https://repositorio.ufrn.br/handle/123456789/46501>. Acesso em: 1 out. 2025.

TONE, K.; TSUTSUI, M. Network DEA: a slacks-based measure approach. **European Journal of Operational Research**, v. 197, n. 1, p. 243-252, 2009. DOI 10.1016/j.ejor.2008.05.027. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0377221708004517>. Acesso em: 18 set. 2025.

TUPY, O.; YAMAGUCHI, L. C. T. Eficiência e produtividade: conceitos e medição. **Agricultura em São Paulo**, v. 45, n. 2, p. 39-51, 1998. Disponível em: <https://www.alice.cnptia.embrapa.br/alice/bitstream/doc/1154700/1/EficienciaProdutividadeConceitosMedicao.pdf>. Acesso em: 9 set. 2025.

VEIDERPASS, A.; MCKELVEY, M. Evaluating the performance of higher education institutions in Europe: a nonparametric efficiency analysis of 944 institutions. **Applied Economics**, v. 48, n. 16, p. 1504-1514, 2016. DOI

10.1080/00036846.2015.1102844. Disponível em:
<https://www.tandfonline.com/doi/abs/10.1080/00036846.2015.1102844>. Acesso em:
27 set. 2025.

VICENTE, J. R. Mudança tecnológica, eficiência e produtividade total dos fatores na agricultura brasileira, 1970-1995. **Economia Aplicada**, v. 8, n. 4, p. 729-760, 2004. DOI 10.11606/1413-8050/ea221373. Disponível em:
https://revistas.usp.br/ecoa/pt_BR/article/view/221373. Acesso em: 30 set. 2025.

WOLSZCZAK-DERLACZ, J. An evaluation and explanation of (in)efficiency in higher education institutions in Europe and the U.S. with the application of two-stage semiparametric DEA. **Research Policy**, v. 46, n. 9, p. 1595-1605, 2017. DOI 10.1016/j.respol.2017.07.010. Disponível em:
<https://www.sciencedirect.com/science/article/pii/S0048733317301221>. Acesso em:
4 out. 2025.

YAMADA, Y.; MATUI, T.; SUGIYAMA, M. New analysis of efficiency based on DEA. **Journal of the Operations Research Society of Japan**, v. 37, n. 2, p. 158-167, 1994. DOI 10.15807/jorsj.37.158. Disponível em:
https://www.jstage.jst.go.jp/article/jorsj/37/2/37_KJ00001202925/_article/-char/ja/. Acesso em: 28 set. 2025.

ZOU, H.; HASTIE, T. Regularization and variable selection via the elastic net. **Journal of the Royal Statistical Society: Series B (Statistical Methodology)**, v. 67, n. 2, p. 301-320, 2005. DOI 10.1111/j.1467-9868.2005.00503.x. Disponível em:
<https://academic.oup.com/jrsssb/article/67/2/301/7109482?guestAccessKey=>. Acesso em: 21 set. 2025.

APÊNDICE A – DMUS E VARIÁVEIS UTILIZADAS

O presente apêndice reúne a base completa de dados utilizada na modelagem da Análise Envoltória de Dados (DEA) e nas análises complementares realizadas ao longo deste estudo. A base contempla todas as Unidades Tomadoras de Decisão (Decision Making Units – DMUs), correspondentes às Instituições de Ensino Superior (IES) públicas federais incluídas no período analisado, bem como o conjunto de variáveis discricionárias e não discricionárias empregadas nos modelos.

Para cada DMU e ano de referência, a base apresenta:

1. Identificação institucional

- Nome da IES, conforme registros oficiais do Inep;
- Código da DMU utilizado nos modelos.

2. Variáveis discricionárias (inputs e outputs)

Representadas pelos indicadores x1, x2, x3, x4, x5, x6, x7, x8 e x9, englobando informações referentes:

- ao corpo docente e técnico-administrativo;
- às matrículas presenciais e a distância;
- à infraestrutura acadêmica física e digital;
- à produção formativa (egressos);
- aos indicadores de qualidade institucional.

Essas variáveis constituem a base operacional da modelagem DEA, refletindo os recursos empregados e os resultados produzidos pelas instituições.

3. Variáveis não discricionárias / ambientais

As variáveis ambientais incluídas no estudo representam fatores estruturais, socioeconômicos e institucionais que influenciam os escores de eficiência, mas que não estão sob controle direto das IES. As variáveis são representadas pelos indicadores z1, z2, z3, z4, z5, z6, z7 e z8, na sequência a seguir:

- PIBPC_MUN (Produto Interno Bruto per capita municipal): aproxima o nível de desenvolvimento socioeconômico do entorno imediato da IES.
- DENS_MUN (Densidade demográfica municipal): indica a concentração populacional, refletindo potenciais pressões sobre a infraestrutura e serviços educacionais.

- TP_REDE (Tipo de rede): dummy que assume valor 0 para instituições públicas e 1 para privadas, captando diferenças relacionadas à governança, fontes de financiamento e mecanismos de accountability.
- ANO_MEC (Idade institucional): ano de credenciamento da IES junto ao MEC, representando seu grau de consolidação acadêmica e administrativa.
- IN_CAPITAL (Sede na capital): dummy que assume valor 0 para instituições sediadas em municípios do interior e 1 para instituições sediadas em capitais estaduais, captando vantagens locacionais.
- LAT (Latitude): coordenada geográfica que indica a posição da sede da IES no eixo norte-sul.
- LON (Longitude): coordenada geográfica que indica a posição da sede da IES no eixo leste-oeste, complementando a latitude.
- DUMMY_FIES (Financiamento estudantil): dummy que assume valor 1 para instituições que possuem alunos financiados pelo Fies e 0 caso contrário, captando a influência de políticas de financiamento estudantil sobre o desempenho e a sustentabilidade institucional.

4. Disponibilização do dataset

Dada a extensão e complexidade da base de dados, optou-se por disponibilizar o arquivo integral em repositório digital permanente, assegurando transparência, reprodutibilidade e acesso público, em conformidade com as boas práticas de ciência aberta. O dataset completo utilizado neste apêndice encontra-se disponível em:

DOI <https://doi.org/10.5281/zenodo.18350199>

O acesso ao material permite a replicação integral dos modelos estimados, bem como a realização de análises adicionais por outros pesquisadores, fortalecendo a robustez metodológica e a utilidade científica deste estudo.

APÊNDICE B – RESULTADOS DE EFICIÊNCIA PARA AS TRÊS ETAPAS

Este apêndice apresenta os resultados consolidados das eficiências estimadas para as Instituições de Ensino Superior (IES) analisadas nas três etapas do modelo de Análise Envoltória de Dados (DEA) em estrutura sequencial. Cada etapa corresponde a um subprocesso produtivo distinto, conforme detalhado no Capítulo 3, permitindo avaliar o desempenho relativo das DMUs de forma desagregada e integrada.

As estimativas foram geradas para dois regimes de retornos à escala — retornos variáveis de escala (VRS) e retornos constantes de escala (CRS) — possibilitando a decomposição da eficiência técnica em suas componentes eficiência técnica pura e eficiência de escala. Também é apresentada a medida de eficiência global, resultante da multiplicação das eficiências técnicas puras das três etapas, sintetizando o desempenho agregado da DMU na cadeia produtiva completa.

A tabela disponibilizada no repositório digital reúne, para cada DMU e ano-base, os seguintes indicadores:

- E. VRS1 – Eficiência técnica com retornos variáveis de escala (VRS) da Etapa 01, sem a adoção do bootstrap.
- E. CRS1 – Eficiência técnica com retornos constantes de escala (CRS) da Etapa 01, sem a adoção do bootstrap.
- EE1 – Eficiência de escala da Etapa 01 (razão entre CRS1 e VRS1).
- E. VRS2 – Eficiência técnica com retornos variáveis de escala (VRS) da Etapa 02, sem a adoção do bootstrap.
- E. CRS2 – Eficiência técnica com retornos constantes de escala (CRS) da Etapa 02, sem a adoção do bootstrap.
- EE2 – Eficiência de escala da Etapa 02 (razão entre CRS2 e VRS2).
- E. VRS3 – Eficiência técnica com retornos variáveis de escala (VRS) da Etapa 03, sem a adoção do bootstrap.
- E. CRS3 – Eficiência técnica com retornos constantes de escala (CRS) da Etapa 03, sem a adoção do bootstrap.
- EE3 – Eficiência de escala da Etapa 03 (razão entre CRS3 e VRS3).
- E.Global – Eficiência Global da DMU, calculada como:

$$E.Global = (E.VRS1) \times (E.VRS2) \times (E.VRS3)$$

Disponibilização dos Resultados

Considerando o volume e o detalhamento da base de resultados, optou-se por disponibilizar integralmente a tabela em repositório digital permanente, assegurando transparência, reprodutibilidade e acesso público. O conjunto completo de resultados das três etapas pode ser consultado em:

DOI <https://doi.org/10.5281/zenodo.18350199>

O acesso ao arquivo permite a reprodução integral dos modelos estimados, bem como a realização de análises adicionais por outros pesquisadores, fortalecendo a robustez e a utilidade científica dos achados.

APÊNDICE C – RESULTADOS DE EFICIÊNCIA PARA AS TRÊS ETAPAS COM BOOTSTRAP

Este apêndice apresenta os resultados das eficiências estimadas para as Instituições de Ensino Superior (IES) nas três etapas do modelo de Análise Envoltória de Dados (DEA), agora com a aplicação do procedimento de bootstrap, conforme a metodologia proposta por Simar e Wilson (2007). O uso do bootstrap permite corrigir o viés inerente às estimativas determinísticas, fornecendo valores mais robustos e estatisticamente consistentes para interpretações inferenciais.

Os resultados aqui disponibilizados abrangem, para cada DMU, as eficiências técnicas puras sob retornos variáveis de escala (VRS) em cada etapa da cadeia produtiva, bem como a eficiência global bootstrap (EGB), obtida pela multiplicação das eficiências parciais ajustadas pelo bootstrap.

Dada a extensão da base de resultados, os dados completos encontram-se disponíveis em repositório digital de acesso aberto, assegurando transparência, reprodutibilidade e interoperabilidade dos achados.

- E. VRS1B – Eficiência técnica com retornos variáveis de escala (VRS) da Etapa 01, com adoção do bootstrap.
- E. VRS2B – Eficiência técnica com retornos variáveis de escala (VRS) da Etapa 02, com adoção do bootstrap.
- E. VRS3B – Eficiência técnica com retornos variáveis de escala (VRS) da Etapa 03, com adoção do bootstrap.
- EGB – Eficiência Global bootstrap, calculada por:
 $EGB = (E.VRS1B) \times (E.VRS2B) \times (E.VRS3B)$

Disponibilização dos Resultados

O conjunto completo dos resultados obtidos com a aplicação do bootstrap encontra-se disponível para consulta no repositório permanente:

DOI <https://doi.org/10.5281/zenodo.18350199>

O arquivo disponibilizado contém os valores estimados para todas as DMUs e para todos os anos analisados, permitindo reprodução integral das análises, além de comparações entre estimativas determinísticas e ajustadas pelo bootstrap, conforme discutido no capítulo metodológico.

APÊNDICE D – RESULTADOS DO IM PARA AS TRÊS ETAPAS

ANO	DMU	IM	TC	PECH	SECH	EC
2020	DMU21095	10,79513	10,37086	1	1,04091	1,04091
2020	DMU4	1,287252	1,166393	1,067087	1,034235	1,103618
2020	DMU4504	1,24236	1,139153	1,058245	1,030575	1,0906
2020	DMU578	1,090902	1,149276	1,004794	0,94468	0,949209
2020	DMU21206	1,087228	1,087228	1	1	1
2020	DMU14724	1,062464	1,184258	1,099988	0,815605	0,897156
2020	DMU5	1,056335	1,130073	1,012888	0,922856	0,93475
2020	DMU4098	1,055115	1,216184	1,061696	0,817147	0,867562
2020	DMU1166	1,048336	1,130251	0,95691	0,969292	0,927525
2020	DMU14408	1,038197	1,130729	1,013188	0,906215	0,918166
2020	DMU594	1,034255	1,143237	0,937023	0,965475	0,904672
2020	DMU693	1,029424	1,130895	1	0,910274	0,910274
2020	DMU3163	1,029352	1,108793	1,068657	0,86871	0,928353
2020	DMU3161	1,029256	0,998489	1,063365	0,969388	1,030813
2020	DMU572	1,026719	1,125638	1	0,912122	0,912122
2020	DMU15497	1,02498	1,121455	0,987341	0,925691	0,913973
2020	DMU633	1,024581	0,934581	1,038902	1,055249	1,096301
2020	DMU548	1,017997	1,127216	1,068657	0,845086	0,903107
2020	DMU592	1,017966	1,130102	1,015584	0,886951	0,900774
2020	DMU18759	1,015767	1,127543	1,000479	0,900437	0,900868
2020	DMU1811	1,014849	1,146823	1	0,884922	0,884922
2020	DMU3984	1,014498	1,128302	0,964873	0,931871	0,899137
2020	DMU4358	1,013457	1,126493	1,058651	0,849814	0,899657
2020	DMU590	1,012925	1,123969	1,028615	0,876133	0,901203
2020	DMU1120	1,011703	1,133181	0,938032	0,951779	0,892799
2020	DMU586	1,011231	1,125579	1,019292	0,881404	0,898409
2020	DMU3188	1,009387	1,118955	1	0,90208	0,90208
2020	DMU1810	1,009055	1,141801	0,999011	0,884614	0,88374
2020	DMU598	1,00896	1,117381	0,982257	0,919279	0,902969
2020	DMU596	1,007969	1,13454	0,959379	0,926056	0,888439
2020	DMU15522	1,007327	1,020633	1,08007	0,913795	0,986963
2020	DMU1808	1,006979	1,145221	1	0,879288	0,879288
2020	DMU602	1,006301	0,964671	1	1,043155	1,043155
2020	DMU3160	1,005727	1,1272	1,018171	0,876311	0,892235
2020	DMU830	1,002672	1,123183	1,009656	0,884168	0,892706
2020	DMU717	1,002469	1,135362	0,97994	0,901025	0,882951
2020	DMU1807	1,002361	1,138272	1	0,880599	0,880599
2020	DMU5322	1,001895	1,133925	0,940431	0,939531	0,883564
2020	DMU1082	1,000298	1,14498	1	0,873638	0,873638
2020	DMU584	0,998605	1,137328	1,006732	0,872156	0,878027
2020	DMU583	0,997529	1,132178	1,017131	0,86623	0,88107
2020	DMU593	0,997443	1,123092	0,975514	0,910415	0,888122
2020	DMU580	0,995294	1,131503	1,037305	0,847988	0,879622
2020	DMU7	0,994044	1,138105	0,956218	0,913411	0,87342
2020	DMU15121	0,993742	1,134224	0,994919	0,880616	0,876142

ANO	DMU	IM	TC	PECH	SECH	EC
2020	DMU575	0,993282	1,131048	0,995798	0,881901	0,878196
2020	DMU574	0,992677	1,125881	0,950555	0,927551	0,881689
2020	DMU6	0,989244	1,127141	0,926784	0,946993	0,877658
2020	DMU694	0,988945	1,136624	1	0,870072	0,870072
2020	DMU18440	0,988642	1,131409	0,9852	0,886942	0,873815
2020	DMU591	0,985736	1,146987	1	0,859414	0,859414
2020	DMU15507	0,984747	1,10988	1,101198	0,805719	0,887256
2020	DMU12	0,981636	1,137402	1	0,863051	0,863051
2020	DMU18812	0,97969	1,127743	0,986287	0,880796	0,868718
2020	DMU1809	0,979373	1,127077	0,935754	0,928608	0,868949
2020	DMU3164	0,979319	1,133961	0,935754	0,92292	0,863626
2020	DMU549	0,979054	1,133215	1,064825	0,811364	0,863961
2020	DMU3183	0,976637	0,977174	1	0,999451	0,999451
2020	DMU3849	0,975884	1,131959	0,996413	0,865223	0,862119
2020	DMU573	0,975275	1,135281	1,044402	0,822538	0,859061
2020	DMU597	0,974979	1,132177	0,976816	0,881593	0,861154
2020	DMU1820	0,974828	1,111669	1,068657	0,820567	0,876905
2020	DMU18506	0,974601	1,130146	1,043258	0,82661	0,862368
2020	DMU577	0,973059	1,133189	0,987593	0,869479	0,858691
2020	DMU3189	0,971748	1,136928	0,885706	0,965008	0,854714
2020	DMU2564	0,970602	1,139941	1	0,851449	0,851449
2020	DMU585	0,970551	1,132407	1	0,857069	0,857069
2020	DMU582	0,970142	1,138107	1	0,852417	0,852417
2020	DMU570	0,970077	1,133704	0,982735	0,870703	0,85567
2020	DMU699	0,969252	1,133276	0,937474	0,912309	0,855266
2020	DMU107	0,968621	1,130487	0,928303	0,922993	0,856818
2020	DMU3162	0,967536	1,141312	0,909859	0,931727	0,847741
2020	DMU634	0,967432	1,13569	0,990118	0,860347	0,851845
2020	DMU2	0,966936	1,128951	1	0,85649	0,85649
2020	DMU15059	0,965318	1,126121	1,028904	0,833126	0,857206
2020	DMU587	0,964881	1,13456	0,990859	0,858291	0,850445
2020	DMU15001	0,963222	1,097733	1,068657	0,821091	0,877465
2020	DMU581	0,962198	1,131684	1	0,850236	0,850236
2020	DMU8	0,95885	1,137645	0,972452	0,866714	0,842838
2020	DMU26	0,95475	0,937985	1,010204	1,007592	1,017873
2020	DMU589	0,954338	1,127062	0,952506	0,888969	0,846748
2020	DMU588	0,954311	1,13236	1	0,842763	0,842763
2020	DMU569	0,954063	1,133695	1	0,841552	0,841552
2020	DMU576	0,951879	1,134718	0,976499	0,859057	0,838868
2020	DMU595	0,948763	1,127859	0,956057	0,879872	0,841207
2020	DMU1578	0,948663	1,12906	0,994153	0,845165	0,840224
2020	DMU789	0,946778	1,132356	0,946547	0,88333	0,836113
2020	DMU4925	0,946443	1,12181	0,997953	0,845406	0,843675
2020	DMU4503	0,943982	1,138482	1,041838	0,795861	0,829158
2020	DMU3	0,936824	1,134668	1	0,825637	0,825637
2020	DMU3279	0,932542	1,126102	1,063921	0,778361	0,828115
2020	DMU15520	0,928218	0,989786	0,920742	1,018523	0,937796

ANO	DMU	IM	TC	PECH	SECH	EC
2020	DMU3165	0,926927	1,134851	0,888528	0,919254	0,816783
2020	DMU4785	0,926673	1,091281	1,065735	0,796784	0,849161
2020	DMU4786	0,924984	1,13214	0,923376	0,884821	0,817023
2020	DMU601	0,923904	1,139511	1,057276	0,766867	0,81079
2020	DMU571	0,917582	1,131199	0,954426	0,849891	0,811159
2020	DMU3184	0,91482	1,062294	0,848137	1,015371	0,861174
2020	DMU5036	0,911377	1,101585	1,097531	0,753812	0,827332
2020	DMU1303	0,910697	1,132008	0,922507	0,872076	0,804496
2020	DMU1812	0,909808	1,091732	0,925448	0,900496	0,833362
2020	DMU579	0,908782	1,135375	1	0,800424	0,800424
2020	DMU1	0,900757	1,139873	1,070793	0,737982	0,790226
2020	DMU599	0,891867	1,100902	1,086081	0,745914	0,810123
2020	DMU16037	0,851352	1,026108	0,853647	0,971935	0,82969
2020	DMU1813	0,844681	1,131926	0,950293	0,785266	0,746233
2020	DMU14509	0,828199	1,140164	0,830026	0,875137	0,726386
2020	DMU17	0,788222	1,135912	1	0,693911	0,693911
2020	DMU4016	0,550236	0,887479	0,615728	1,006936	0,619999
2021	DMU4016	1,533468	0,978435	1,610488	0,973161	1,567265
2021	DMU15520	1,266758	0,884039	1,160648	1,234587	1,432921
2021	DMU17	1,22959	0,867448	1	1,417479	1,417479
2021	DMU14509	1,19656	0,954639	1,36545	0,917951	1,253416
2021	DMU3188	1,157515	0,868107	1	1,333377	1,333377
2021	DMU1812	1,150842	0,932369	1,150132	1,073199	1,234321
2021	DMU15507	1,149324	0,922957	1,136284	1,095908	1,245263
2021	DMU16037	1,130787	0,938203	1,171444	1,028875	1,205269
2021	DMU3	1,120351	0,852215	1	1,314634	1,314634
2021	DMU599	1,096212	0,918519	1,068657	1,116781	1,193456
2021	DMU579	1,093657	0,868322	1	1,259507	1,259507
2021	DMU3184	1,08062	0,902989	1,090052	1,09785	1,196714
2021	DMU1813	1,080462	0,89027	1,086081	1,117444	1,213635
2021	DMU1820	1,064202	0,881905	1	1,206708	1,206708
2021	DMU601	1,06344	0,869023	0,917207	1,33418	1,22372
2021	DMU3165	1,056368	0,877616	1,084252	1,110146	1,203678
2021	DMU15001	1,053069	0,904672	1	1,164034	1,164034
2021	DMU789	1,046918	0,829493	1,096514	1,151028	1,262118
2021	DMU3161	1,046404	0,985261	0,935754	1,134975	1,062057
2021	DMU5036	1,043402	0,89992	0,862095	1,344907	1,159439
2021	DMU1303	1,04225	0,857879	1,081373	1,123493	1,214916
2021	DMU1	1,041724	0,861131	1	1,209716	1,209716
2021	DMU1578	1,033245	0,873917	1,000469	1,181761	1,182315
2021	DMU4925	1,030509	0,830697	1,040803	1,191902	1,240536
2021	DMU576	1,028632	0,857876	1,061492	1,129584	1,199045
2021	DMU12	1,02812	0,871183	0,952514	1,238977	1,180143
2021	DMU3183	1,026362	1,015057	1	1,011137	1,011137
2021	DMU1809	1,024854	0,880131	0,93849	1,240752	1,164434
2021	DMU5322	1,023126	0,866544	0,996416	1,184944	1,180697
2021	DMU1808	1,017242	0,873523	0,935754	1,24448	1,164527

ANO	DMU	IM	TC	PECH	SECH	EC
2021	DMU694	1,017178	0,873355	1	1,164679	1,164679
2021	DMU3163	1,016504	0,887999	0,935754	1,223305	1,144712
2021	DMU4785	1,015865	0,888253	1	1,143666	1,143666
2021	DMU3279	1,015743	0,86852	0,937269	1,247785	1,16951
2021	DMU591	1,01514	0,875099	1	1,160029	1,160029
2021	DMU585	1,014402	0,858367	1	1,18178	1,18178
2021	DMU570	1,012538	0,867018	1,036703	1,126493	1,167839
2021	DMU3189	1,007149	0,87621	1,048409	1,096364	1,149437
2021	DMU597	1,005902	0,893231	1,013546	1,111088	1,126139
2021	DMU4358	1,004146	0,857963	0,992413	1,179332	1,170384
2021	DMU8	0,999275	0,88476	0,986766	1,144579	1,129431
2021	DMU582	0,999192	0,872793	1	1,144821	1,144821
2021	DMU693	0,9988	0,880523	1	1,134326	1,134326
2021	DMU1810	0,998403	0,875123	1,063621	1,07263	1,140872
2021	DMU4503	0,997079	0,849501	1	1,173723	1,173723
2021	DMU573	0,99669	0,869978	1	1,145649	1,145649
2021	DMU602	0,996518	1,026545	1	0,970749	0,970749
2021	DMU592	0,995599	0,839413	1,027799	1,153985	1,186066
2021	DMU581	0,994936	0,856091	1	1,162185	1,162185
2021	DMU15121	0,992251	0,857318	1,002214	1,154833	1,15739
2021	DMU584	0,990995	0,879782	1,003642	1,122322	1,12641
2021	DMU15497	0,989542	0,843949	1,061059	1,105041	1,172514
2021	DMU4786	0,989107	0,875843	1,067697	1,057716	1,12932
2021	DMU583	0,988442	0,865822	1,018324	1,121078	1,141621
2021	DMU6	0,987882	0,8437	1,049307	1,115872	1,170892
2021	DMU575	0,986666	0,861863	1,030857	1,110537	1,144806
2021	DMU14408	0,986583	0,862781	1,046281	1,09291	1,143491
2021	DMU569	0,984522	0,866131	1	1,136689	1,136689
2021	DMU26	0,983846	1,018495	0,992743	0,973041	0,96598
2021	DMU633	0,98356	0,98356	1	1	1
2021	DMU18440	0,983371	0,837833	1,054313	1,113244	1,173707
2021	DMU2	0,983145	0,848075	1	1,159267	1,159267
2021	DMU1811	0,980346	0,862888	1	1,136122	1,136122
2021	DMU588	0,979427	0,83571	1	1,17197	1,17197
2021	DMU594	0,977616	0,862722	1,008298	1,12385	1,133177
2021	DMU580	0,976618	0,86328	0,981234	1,152923	1,131288
2021	DMU3162	0,97541	0,875445	1,137987	0,979086	1,114187
2021	DMU107	0,973671	0,80895	1,058067	1,137569	1,203624
2021	DMU595	0,973136	0,835222	1,045963	1,113923	1,165122
2021	DMU699	0,969572	0,836093	1,014843	1,142685	1,159646
2021	DMU598	0,967757	0,797702	1,043538	1,162565	1,213181
2021	DMU5	0,964621	0,853544	1	1,130136	1,130136
2021	DMU571	0,960087	0,869422	0,980436	1,126317	1,104282
2021	DMU7	0,960036	0,863373	0,963161	1,15449	1,11196
2021	DMU1082	0,957363	0,877506	1	1,091004	1,091004
2021	DMU578	0,955967	0,877491	1,023892	1,064012	1,089432
2021	DMU634	0,954256	0,851826	0,981234	1,141672	1,120247

ANO	DMU	IM	TC	PECH	SECH	EC
2021	DMU586	0,953686	0,837994	1	1,138057	1,138057
2021	DMU593	0,953499	0,828286	1,072138	1,073715	1,151172
2021	DMU1120	0,953486	0,871707	1,066062	1,026033	1,093815
2021	DMU18812	0,952564	0,807507	1,017152	1,159744	1,179636
2021	DMU549	0,947918	0,865018	0,937226	1,169232	1,095836
2021	DMU574	0,943975	0,848205	0,984429	1,130513	1,112909
2021	DMU572	0,943871	0,839343	1	1,124535	1,124535
2021	DMU590	0,942588	0,838704	0,967498	1,161617	1,123862
2021	DMU1807	0,942231	0,86443	1	1,090003	1,090003
2021	DMU548	0,941222	0,839702	1	1,120901	1,120901
2021	DMU15522	0,937607	1,034896	0,980744	0,92378	0,905992
2021	DMU577	0,937409	0,865675	0,989956	1,093852	1,082865
2021	DMU18506	0,930611	0,830476	0,987121	1,135195	1,120574
2021	DMU596	0,918474	0,855246	0,975375	1,101043	1,07393
2021	DMU587	0,914863	0,877381	1	1,04272	1,04272
2021	DMU717	0,910028	0,833291	0,97845	1,116142	1,092089
2021	DMU2564	0,909006	0,871138	0,855966	1,219055	1,043469
2021	DMU3849	0,902588	0,846309	0,992286	1,07479	1,066499
2021	DMU1166	0,899598	0,873934	1,045031	0,98501	1,029366
2021	DMU15059	0,89572	0,893546	1	1,002433	1,002433
2021	DMU18759	0,89569	0,862954	0,973779	1,065883	1,037934
2021	DMU3160	0,888981	0,883694	0,970699	1,036349	1,005983
2021	DMU3984	0,884076	0,83166	1,01859	1,043625	1,063026
2021	DMU830	0,871542	0,821869	1,051724	1,008286	1,060439
2021	DMU4098	0,868908	0,752358	0,935754	1,234207	1,154914
2021	DMU4504	0,853662	0,895609	1	0,953163	0,953163
2021	DMU14724	0,850149	0,773889	0,893442	1,229561	1,098542
2021	DMU3164	0,839093	0,900683	1	0,931618	0,931618
2021	DMU4	0,811285	0,878846	1	0,923125	0,923125
2021	DMU21206	0,764313	0,764313	1	1	1
2021	DMU589	0,745913	0,947319	1,049863	0,749997	0,787394
2021	DMU21095	0,064574	0,064574	1	1	1
2022	DMU589	1,722796	1,109913	1	1,55219	1,55219
2022	DMU3164	1,541945	1,07652	1,068657	1,34032	1,432342
2022	DMU21095	1,514637	1,514637	1	1	1
2022	DMU21206	1,505527	1,505527	1	1	1
2022	DMU15522	1,366021	1,063935	0,975228	1,316546	1,283933
2022	DMU15059	1,158024	1,054057	1	1,098635	1,098635
2022	DMU18506	1,152462	1,094	1,033549	1,019245	1,053439
2022	DMU3160	1,151517	1,050814	1,002709	1,092872	1,095833
2022	DMU717	1,146149	1,121256	1	1,022201	1,022201
2022	DMU4785	1,136909	1,055226	1,011716	1,064931	1,077408
2022	DMU830	1,133919	1,037318	0,935754	1,168176	1,093126
2022	DMU571	1,13255	1,044911	1,068657	1,014237	1,083871
2022	DMU3849	1,115612	1,057606	1,003662	1,050998	1,054847
2022	DMU1578	1,106274	1,016226	1,042877	1,043854	1,088611
2022	DMU1166	1,099153	1,06832	1	1,028861	1,028861

ANO	DMU	IM	TC	PECH	SECH	EC
2022	DMU549	1,094922	1,063331	1,0139	1,015593	1,02971
2022	DMU1	1,091001	1,081143	1	1,009118	1,009118
2022	DMU14509	1,08844	1,066659	0,878917	1,160997	1,020421
2022	DMU596	1,087189	1,105398	1,018718	0,965455	0,983527
2022	DMU4503	1,08452	1,10451	1	0,981902	0,981902
2022	DMU4016	1,079914	0,998676	1,013331	1,06712	1,081346
2022	DMU588	1,071777	1,075709	1	0,996345	0,996345
2022	DMU574	1,062372	1,062261	1,01904	0,981418	1,000104
2022	DMU3984	1,061333	1,066245	0,970419	1,025736	0,995394
2022	DMU3183	1,060031	1,076672	1	0,984543	0,984543
2022	DMU3162	1,059648	1,094623	0,951889	1,016976	0,968048
2022	DMU1082	1,05946	1,090197	1	0,971806	0,971806
2022	DMU548	1,059376	1,025871	1	1,03266	1,03266
2022	DMU7	1,058457	1,089245	1,041853	0,932699	0,971735
2022	DMU18759	1,057209	1,046686	0,936612	1,078413	1,010054
2022	DMU599	1,056669	1,086695	0,861588	1,128579	0,972369
2022	DMU601	1,053659	1,083192	1,090266	0,8922	0,972735
2022	DMU569	1,053654	1,027103	1	1,02585	1,02585
2022	DMU634	1,05297	1,098816	1,018282	0,941072	0,958276
2022	DMU595	1,052098	1,091447	1	0,963948	0,963948
2022	DMU18440	1,046915	1,082131	1	0,967457	0,967457
2022	DMU577	1,046215	1,054341	1	0,992292	0,992292
2022	DMU699	1,044572	1,124645	1,094045	0,84896	0,928801
2022	DMU789	1,040556	1,103719	0,98046	0,961561	0,942773
2022	DMU4098	1,039732	1,108685	1,006561	0,931693	0,937806
2022	DMU2564	1,038826	1,088747	1,028379	0,927818	0,954148
2022	DMU693	1,035506	1,055981	1	0,980611	0,980611
2022	DMU602	1,034685	1,034071	1	1,000594	1,000594
2022	DMU8	1,0345	1,070733	1,030781	0,93731	0,966161
2022	DMU590	1,032911	1,042883	1,062392	0,932272	0,990439
2022	DMU4786	1,030033	1,060531	0,920742	1,054849	0,971243
2022	DMU5036	1,029842	1,078321	1,085441	0,879865	0,955042
2022	DMU1820	1,029205	1,084567	0,935754	1,014106	0,948954
2022	DMU572	1,027647	1,030197	1	0,997524	0,997524
2022	DMU1811	1,027596	1,089336	1	0,943323	0,943323
2022	DMU5322	1,026704	1,082877	1,010458	0,938313	0,948126
2022	DMU573	1,024668	1,058223	1	0,968292	0,968292
2022	DMU14724	1,024117	1,086602	1,010051	0,933116	0,942495
2022	DMU597	1,023827	1,02203	1,019615	0,982487	1,001759
2022	DMU4358	1,023693	1,040474	1,055421	0,932209	0,983872
2022	DMU1813	1,023514	1,095566	0,920742	1,014652	0,934233
2022	DMU576	1,02346	1,072671	1	0,954123	0,954123
2022	DMU1807	1,022817	1,093962	1	0,934966	0,934966
2022	DMU582	1,022108	1,067269	1	0,957685	0,957685
2022	DMU3189	1,021248	1,072466	1,025373	0,92868	0,952243
2022	DMU586	1,02064	1,036395	0,993691	0,991051	0,984799
2022	DMU1120	1,0206	1,0951	1	0,93197	0,93197

ANO	DMU	IM	TC	PECH	SECH	EC
2022	DMU3165	1,020203	1,060653	1,019944	0,943055	0,961863
2022	DMU14408	1,01785	1,064792	0,968108	0,987405	0,955915
2022	DMU581	1,017729	1,039868	1	0,97871	0,97871
2022	DMU15497	1,016403	1,043708	0,988457	0,985211	0,973838
2022	DMU584	1,015086	1,071618	0,996007	0,951044	0,947247
2022	DMU2	1,013431	1,030708	0,985455	0,99775	0,983238
2022	DMU16037	1,013064	1,013064	1	1	1
2022	DMU26	1,012868	1,05055	0,899023	1,072421	0,964131
2022	DMU583	1,01086	1,041992	0,988626	0,981284	0,970123
2022	DMU694	1,008447	1,074987	1	0,938101	0,938101
2022	DMU591	1,008284	1,073014	1	0,939675	0,939675
2022	DMU575	1,008161	1,035036	1	0,974035	0,974035
2022	DMU1810	1,007855	1,082653	1	0,930912	0,930912
2022	DMU570	1,005939	1,033048	1,004427	0,969466	0,973758
2022	DMU15121	1,003816	1,099868	0,995571	0,91673	0,91267
2022	DMU1808	1,002438	1,086123	1	0,922951	0,922951
2022	DMU580	1,001089	1,039483	0,979811	0,982909	0,963065
2022	DMU107	1,000327	1,095127	0,984621	0,927702	0,913435
2022	DMU4	0,998608	1,029019	0,939886	1,032516	0,970447
2022	DMU585	0,996598	1,045388	1	0,953328	0,953328
2022	DMU12	0,996582	1,091408	0,990289	0,922069	0,913115
2022	DMU594	0,994097	1,083267	1,058426	0,867028	0,917685
2022	DMU3279	0,994097	1,095427	0,998384	0,908966	0,907497
2022	DMU1303	0,9939	1,089427	1,068657	0,853702	0,912314
2022	DMU18812	0,992404	1,084321	0,933895	0,980014	0,915231
2022	DMU578	0,992261	1,034491	0,960382	0,998746	0,959178
2022	DMU592	0,991838	1,085662	0,971886	0,940006	0,913578
2022	DMU587	0,991358	1,065488	1	0,930426	0,930426
2022	DMU4504	0,990783	1,07959	1	0,91774	0,91774
2022	DMU15507	0,989961	1,089354	1	0,90876	0,90876
2022	DMU4925	0,97768	1,037082	1	0,942721	0,942721
2022	DMU5	0,972893	1,063449	0,973428	0,939819	0,914846
2022	DMU3	0,970796	1,075677	1	0,902497	0,902497
2022	DMU6	0,96749	1,062814	1,028297	0,885259	0,91031
2022	DMU15001	0,966738	1,053995	0,935754	0,980186	0,917213
2022	DMU1812	0,95517	1,012665	0,935754	1,007984	0,943225
2022	DMU1809	0,953869	1,086978	0,981088	0,894458	0,877542
2022	DMU17	0,940133	1,062921	1	0,88448	0,88448
2022	DMU3161	0,934252	1,084963	1	0,861091	0,861091
2022	DMU579	0,910787	1,051323	1	0,866325	0,866325
2022	DMU598	0,90993	0,998344	0,923772	0,98665	0,911439
2022	DMU593	0,907476	1,04604	1,068657	0,811799	0,867535
2022	DMU3188	0,906481	1,059138	0,935754	0,914628	0,855867
2022	DMU3184	0,905481	1,047751	0,969722	0,891197	0,864214
2022	DMU633	0,903155	0,903155	1	1	1
2022	DMU15520	0,901718	1,253276	0,794155	0,90598	0,719489
2022	DMU3163	0,899008	1,07555	1	0,835859	0,835859

ANO	DMU	IM	TC	PECH	SECH	EC
2023	DMU16037	6,53993	6,53993	1	1	1
2023	DMU21206	1,33413	1,33413	1	1	1
2023	DMU21095	1,281104	1,281104	1	1	1
2023	DMU587	1,239373	0,939061	1,068657	1,235009	1,3198
2023	DMU593	1,174685	1,009452	0,964996	1,205896	1,163686
2023	DMU4786	1,159944	0,981747	1,133884	1,042003	1,18151
2023	DMU1303	1,116209	0,956355	0,933126	1,250795	1,167149
2023	DMU3184	1,102114	0,955903	1,101266	1,046937	1,152956
2023	DMU1809	1,087657	0,928465	1,090979	1,073767	1,171458
2023	DMU1813	1,086851	0,933983	1,09016	1,067433	1,163673
2023	DMU17	1,08518	0,953406	1	1,138213	1,138213
2023	DMU633	1,072548	1,178078	0,982728	0,926423	0,910421
2023	DMU4785	1,068115	0,963947	1,147207	0,96588	1,108064
2023	DMU2564	1,064557	0,941718	1,081994	1,044776	1,130441
2023	DMU579	1,050572	0,973947	1	1,078675	1,078675
2023	DMU15001	1,049	0,948807	1,002674	1,102651	1,105599
2023	DMU15507	1,046554	0,930881	0,92166	1,219822	1,124261
2023	DMU4016	1,040772	1,027274	1,00903	1,004074	1,01314
2023	DMU1807	1,039774	0,933937	0,920052	1,210065	1,113323
2023	DMU1820	1,036953	0,930291	1,068657	1,043042	1,114654
2023	DMU3161	1,036679	0,913772	1,068657	1,061618	1,134506
2023	DMU18812	1,030963	0,986668	1,035969	1,008615	1,044894
2023	DMU3849	1,030618	0,972459	1,041115	1,017953	1,059807
2023	DMU15059	1,027381	0,998465	1	1,02896	1,02896
2023	DMU26	1,026325	0,984706	1,04882	0,993751	1,042265
2023	DMU107	1,022147	0,972497	1,051892	0,999204	1,051055
2023	DMU3279	1,019417	0,942258	0,963853	1,122462	1,081888
2023	DMU4098	1,017571	0,927679	1,022201	1,073076	1,096899
2023	DMU18759	1,017439	0,977715	1,080104	0,963453	1,04063
2023	DMU15121	1,015052	0,936555	1,013904	1,068952	1,083815
2023	DMU577	1,010521	0,966272	1,017135	1,028176	1,045794
2023	DMU18440	1,009112	0,992718	1	1,016514	1,016514
2023	DMU8	1,007376	0,943393	1,012368	1,054777	1,067823
2023	DMU14724	1,006808	0,933815	1,054046	1,022883	1,078166
2023	DMU15520	1,006434	0,925959	1,017463	1,068255	1,08691
2023	DMU2	1,004814	1,011942	1,01476	0,978513	0,992956
2023	DMU18506	1,003712	0,938669	1,025359	1,042847	1,069293
2023	DMU602	1,002423	1,051914	1	0,952951	0,952951
2023	DMU3162	1,001814	0,933766	1,009398	1,062885	1,072874
2023	DMU1120	0,999348	0,919735	1	1,086561	1,086561
2023	DMU1082	0,998925	0,929028	1	1,075236	1,075236
2023	DMU3984	0,996276	0,977952	1,020599	0,998175	1,018737
2023	DMU594	0,995083	0,933977	1	1,065426	1,065426
2023	DMU1808	0,993218	0,931734	1,068657	0,997503	1,065988
2023	DMU14408	0,992808	0,950817	1,04603	0,998216	1,044163
2023	DMU583	0,992441	0,978739	1,015225	0,998792	1,013999
2023	DMU591	0,990474	0,948046	1	1,044753	1,044753

ANO	DMU	IM	TC	PECH	SECH	EC
2023	DMU548	0,989391	1,021964	1	0,968128	0,968128
2023	DMU596	0,988093	0,952313	1,004626	1,032795	1,037572
2023	DMU3163	0,988029	0,956073	0,920742	1,122382	1,033424
2023	DMU1811	0,987311	0,927497	1,006883	1,057214	1,06449
2023	DMU578	0,986191	0,994807	1,011158	0,9804	0,991339
2023	DMU717	0,98514	0,951346	1,068657	0,968995	1,035523
2023	DMU830	0,984697	1,003669	1,068657	0,918065	0,981097
2023	DMU1166	0,984215	0,934241	1,007003	1,046165	1,053491
2023	DMU4925	0,983624	1,018305	0,960706	1,005451	0,965943
2023	DMU575	0,983215	0,996974	1	0,986199	0,986199
2023	DMU3189	0,983193	0,939958	1,050262	0,995939	1,045997
2023	DMU3183	0,982905	0,928064	1	1,059092	1,059092
2023	DMU699	0,982678	0,93234	0,953371	1,105542	1,053992
2023	DMU3165	0,982167	0,95112	1,057145	0,976823	1,032643
2023	DMU569	0,98125	1,015681	1	0,9661	0,9661
2023	DMU590	0,981039	1,01538	0,92409	1,045546	0,966179
2023	DMU598	0,980386	1,008349	0,995385	0,976777	0,972269
2023	DMU573	0,979713	0,956326	1	1,024455	1,024455
2023	DMU595	0,979645	0,957264	1	1,02338	1,02338
2023	DMU14509	0,979058	0,931742	1,137763	0,92355	1,050782
2023	DMU584	0,978614	0,945487	1,00503	1,029857	1,035037
2023	DMU4358	0,978497	0,981876	0,93099	1,070429	0,996559
2023	DMU12	0,977534	0,936494	1,000458	1,043345	1,043823
2023	DMU6	0,976522	0,96639	0,924161	1,093407	1,010485
2023	DMU694	0,976513	0,961528	1	1,015584	1,015584
2023	DMU4	0,975888	0,999518	0,995604	0,980669	0,976358
2023	DMU5	0,975847	0,967938	1,027298	0,981381	1,00817
2023	DMU597	0,975386	0,961546	0,999774	1,014623	1,014393
2023	DMU582	0,975203	0,938524	1	1,039082	1,039082
2023	DMU7	0,974427	0,947728	0,99822	1,030005	1,028171
2023	DMU586	0,974378	1,023598	1,006349	0,945909	0,951915
2023	DMU3	0,973881	0,961619	1	1,012752	1,012752
2023	DMU1810	0,973431	0,933817	1	1,042422	1,042422
2023	DMU572	0,972646	1,020674	0,990593	0,961994	0,952944
2023	DMU576	0,972151	0,948699	1	1,02472	1,02472
2023	DMU588	0,97183	0,980027	1	0,991636	0,991636
2023	DMU592	0,97172	0,956386	1,000821	1,015201	1,016034
2023	DMU585	0,971102	0,968992	1	1,002177	1,002177
2023	DMU634	0,970615	0,936498	1,04947	0,987574	1,03643
2023	DMU570	0,968909	0,983885	1	0,984779	0,984779
2023	DMU581	0,967687	0,974167	1	0,993348	0,993348
2023	DMU1578	0,965616	0,995871	0,986102	0,983285	0,969619
2023	DMU4504	0,9642	0,934303	1	1,031999	1,031999
2023	DMU3160	0,962817	0,958831	0,956529	1,049792	1,004157
2023	DMU3188	0,96061	0,962182	1	0,998366	0,998366
2023	DMU599	0,959497	0,927671	1,086081	0,952331	1,034308
2023	DMU574	0,959311	0,969887	0,985594	1,003553	0,989095

ANO	DMU	IM	TC	PECH	SECH	EC
2023	DMU5036	0,958492	0,912767	1,068657	0,982631	1,050096
2023	DMU580	0,955252	0,987476	0,978397	0,988727	0,967367
2023	DMU15497	0,954366	1,002923	0,954366	0,997085	0,951584
2023	DMU693	0,952806	0,976728	1	0,975508	0,975508
2023	DMU789	0,952646	0,953509	1,00062	0,998475	0,999094
2023	DMU1	0,94882	0,942412	1	1,0068	1,0068
2023	DMU5322	0,947725	0,932396	1,056122	0,962427	1,01644
2023	DMU15522	0,947247	0,94967	1,041958	0,957283	0,997449
2023	DMU601	0,945495	0,929696	1	1,016994	1,016994
2023	DMU589	0,941991	1,009786	1	0,932862	0,932862
2023	DMU1812	0,933703	0,951938	1	0,980844	0,980844
2023	DMU4503	0,927588	0,953338	1	0,972989	0,972989
2023	DMU549	0,922454	0,960601	0,957787	1,002612	0,960289
2023	DMU571	0,890189	0,970128	0,935754	0,980599	0,9176
2023	DMU3164	0,410243	0,966135	0,773867	0,548703	0,424623

Legenda	
IM	Índice de Malmquist de Produtividade Total: valor global que indica a variação da produtividade de uma unidade (DMU) entre dois períodos consecutivos. Valores superiores a 1 indicam ganho de produtividade, enquanto valores inferiores a 1 indicam perda de produtividade.
TC	Technical Change – Mudança Tecnológica: representa o deslocamento da fronteira de produção ao longo do tempo. Valores acima de 1 indicam avanço tecnológico, enquanto valores abaixo de 1 sugerem retrocesso tecnológico.
PECH	Pure Efficiency Change – Mudança de Eficiência Técnica Pura: reflete alterações na eficiência relativa das DMUs em comparação à fronteira de produção, desconsiderando o efeito de escala. Valores maiores que 1 representam melhorias na gestão ou no uso dos recursos.
SECH	Scale Efficiency Change – Mudança de Eficiência de Escala: capta as variações na eficiência associadas ao porte da operação da DMU. Valores acima de 1 indicam que a unidade se aproximou da escala ótima de produção.
EC	Efficiency Change – Mudança de Eficiência Técnica: resultado da multiplicação entre PECH e SECH ($EC=PECH \times SECH$), indicando a variação global da eficiência técnica entre os períodos.

APÊNDICE E – SCRIPT UTILIZADO NO SOFTWARE R-STUDIO

```

# =====
# APÊNDICE E – SCRIPT UTILIZADO NO SOFTWARE RSTUDIO
# =====
# Ambiente computacional utilizado na análise:
#
# Software: RStudio
# Versão: 2026.01.1+403
# Data de lançamento: 19 fev. 2026
# Tamanho do instalador: 314.19 MB
# Hash de verificação (SHA-256): 84442594
#
# Linguagem utilizada: R
# Plataforma de execução: RStudio Desktop
#
# O registro da versão do ambiente de desenvolvimento visa
# assegurar a rastreabilidade e a reprodutibilidade das rotinas
# computacionais empregadas no tratamento dos dados, construção
# das bases analíticas e estimação dos modelos utilizados nesta
# pesquisa.
#
# =====
# 1. TRATAMENTO E CONSOLIDAÇÃO DOS DADOS IES / CURSOS
# =====

# Inicialmente, procede-se ao carregamento dos pacotes requeridos
# para as etapas de importação, tratamento, transformação e futura
# modelagem dos dados. Adota-se instalação condicional, de modo que
# bibliotecas ausentes no ambiente local sejam automaticamente
# instaladas antes do carregamento, o que favorece a portabilidade
# e a reprodutibilidade da rotina computacional.

pkgs1 <- c("dear", "readr", "dplyr", "readxl", "writexl", "stringr", "zoo")
to_install_1 <- pkgs1[!pkgs1 %in% rownames(installed.packages())]

if (length(to_install_1)) install.packages(to_install_1, dependencies = TRUE)
invisible(lapply(pkgs1, library, character.only = TRUE))

# Define-se um diretório único de trabalho, no qual permanecem
# centralizados os arquivos de entrada, os objetos processados e
# os produtos intermediários e finais gerados ao longo do script.
# Essa padronização reduz a dispersão de arquivos, simplifica a
# manutenção da rotina e assegura maior rastreabilidade dos outputs.

base_dir <- "E:/Mestrado_Daniel/00 DISSERTAÇÃO_FINAL/05 APÊNDICES/APÊNDICE E – Script
utilizado no software R-STUDIO"
setwd(base_dir)

# Em seguida, realiza-se a importação das bases consolidadas de
# microdados do Censo da Educação Superior, referentes ao cadastro
# de Instituições de Ensino Superior (IES) e ao cadastro de cursos,
# abrangendo o período de 2013 a 2023.
#
# A base de IES é lida com separador ponto e vírgula, respeitando
# a estrutura original do arquivo exportado. Já a base de cursos é
# importada com tratamento explícito da codificação UTF-8, a fim de
# preservar corretamente caracteres especiais e evitar inconsistências
# textuais em variáveis nominais.

```

```

dados_ies_2013_2023 <- read_delim(
  file = file.path(base_dir, "02_MICRODADOS_CADASTRO_IES_2013_2023.csv"),
  delim = ";",
  escape_double = FALSE,
  trim_ws = TRUE
)

dados_cursos_2013_2023 <- read_csv(
  file = file.path(base_dir, "02_MICRODADOS_CADASTRO_CURSOS_2013_2023.csv"),
  locale = locale(encoding = "UTF-8"),
  show_col_types = FALSE
)

# Na base de IES, procede-se ao recorte analítico da amostra,
# restringindo-se a observação às categorias administrativas de
# interesse do estudo. Na sequência, é construída a variável
# indicadora de rede institucional (TP_REDE), empregada para
# distinguir as instituições públicas das privadas.
#
# Convenciona-se:
# 1 = pública
# 2 = privada
#
# Essa variável será utilizada nas etapas subsequentes de análise
# comparativa da eficiência entre os dois segmentos institucionais.

dados_ies_2013_2023 <- dados_ies_2013_2023 %>%
  filter(TP_CATEGORIA_ADMINISTRATIVA %in% c(1, 4, 5, 6, 8, 9)) %>%
  mutate(
    TP_REDE = if_else(TP_CATEGORIA_ADMINISTRATIVA == 1, 1, 2)
  )

# Posteriormente, são construídas proxies institucionais associadas
# à infraestrutura informacional e digital das IES. Tais medidas
# sintéticas buscam representar, de forma operacionalizável, a
# disponibilidade de serviços acadêmico-informacionais e a dimensão
# do acervo eletrônico institucional.
#
# Para isso, selecionam-se variáveis binárias relacionadas ao acesso
# a bases de dados, repositórios, serviços digitais e ferramentas de
# busca e interação, cuja soma resulta em um indicador agregado de
# serviços oferecidos.
#
# Adicionalmente, a soma das quantidades de periódicos eletrônicos
# e livros eletrônicos é utilizada como proxy da disponibilidade de
# acervo digital.
#
# Em ambos os casos, emprega-se tratamento robusto a valores ausentes
# (na.rm = TRUE), evitando que observações com dados faltantes sejam
# indevidamente convertidas em valores perdidos nas variáveis-síntese.

colunas_servicos <- c(
  "IN_ACESSO_PORTAL_CAPES",
  "IN_ACESSO_OUTRAS_BASES",
  "IN_ASSINA_OUTRA_BASE",
  "IN_REPOSITARIO_INSTITUCIONAL",
  "IN_BUSCA_INTEGRADA",
  "IN_SERVICO_INTERNET",
  "IN_PARTICIPA_REDE_SOCIAL",
  "IN_CATALOGO_ONLINE"
)

```

```

)

dados_ies_2013_2023 <- dados_ies_2013_2023 %>%
  mutate(
    SOMATORIO_SERVICOS = rowSums(across(all_of(colunas_servicos)), na.rm = TRUE),
    SOMATORIO_ACERVO = rowSums(across(c(QT_PERIODICO_ELETRONICO,
    QT_LIVRO_ELETRONICO)), na.rm = TRUE)
  )

# Na base de cursos, constrói-se o indicador de abandono a partir da
# agregação das situações de desvinculação e falecimento. Essa opção
# metodológica permite sintetizar, em nível de curso, eventos de saída
# do estudante que reduzem a permanência até a conclusão.
#
# O indicador resultante (QT_ABANDONO) será posteriormente agregado ao
# nível institucional, compondo uma medida de resultado indesejado no
# desempenho das IES ao longo do tempo.

colunas_abandono <- c("QT_SIT_DESVINCULADO", "QT_SIT_FALECIDO")

dados_cursos_2013_2023 <- dados_cursos_2013_2023 %>%
  mutate(
    QT_ABANDONO = rowSums(select(., all_of(colunas_abandono)), na.rm = TRUE)
  )

# Por fim, realiza-se a seleção das variáveis essenciais da base de IES,
# preservando apenas os campos necessários à identificação das unidades
# analisadas, à caracterização institucional, à mensuração dos insumos,
# às proxies de infraestrutura informacional e à localização geográfica.
#
# Essa etapa reduz a dimensionalidade da base, melhora a organização do
# banco analítico e contribui para maior eficiência nas operações de
# junção, agregação e modelagem subsequentes.

dados_selecionados_ies <- dados_ies_2013_2023 %>%
  select(
    NU_ANO_CENSO, CO_IES, NO_IES, SG_IES,
    TP_ORGANIZACAO_ACADEMICA, TP_REDE, TP_CATEGORIA_ADMINISTRATIVA,
    QT_DOC_EXE, QT_TEC_TOTAL,
    SOMATORIO_SERVICOS, SOMATORIO_ACERVO,
    CO_MUNICIPIO_IES, NO_MUNICIPIO_IES, CO_MESORREGIAO_IES,
    NO_MESORREGIAO_IES
  )
# =====
# 1.4. ENRIQUECIMENTO DA BASE: MUNICÍPIOS E IGC
# =====

# Objetivo:
# Enriquecer a base de DMUs no nível IES-ano com:
# (i) informação espacial, a partir do centroide do município de
# localização da IES, permitindo derivação de variáveis locacionais;
# e (ii) indicador qualitativo institucional (IGC contínuo),
# posteriormente utilizado tanto na modelagem quanto nas análises
# explicativas dos diferenciais de eficiência.

# -----
# (i) Enriquecimento espacial: associação do município da IES
# ao respectivo centroide geográfico municipal
# -----

municipios <- read_csv(

```

```

file = file.path(base_dir, "br_bd_diretorios_brasil_municipio.csv.gz"),
locale = locale(encoding = "UTF-8"),
show_col_types = FALSE
)
dados_selecionados_ies <- dados_selecionados_ies %>%
  left_join(
    municipios %>%
      select(id_municipio, centroide),
    by = c("CO_MUNICIPIO_IES" = "id_municipio")
  )
# -----
# (ii) Incorporação do Indicador Geral de Cursos (IGC contínuo)
# -----
# O IGC contínuo é tratado como proxy de qualidade institucional,
# sendo integrado ao painel no nível IES-ano. A preparação da base
# envolve padronização dos nomes das variáveis, expansão explícita
# do painel e imputação de lacunas temporais.

igc_2013_2023 <- read_excel(
  path = file.path(base_dir, "igc_2013_2023.xlsx")
) %>%
  rename(
    NU_ANO_CENSO = Ano,
    CO_IES = `Cód.IES`,
    IGC_CONTINUO = `IGC (Contínuo)`
  )
# Expansão do painel IES x Ano, incluindo explicitamente 2020,
# a fim de assegurar continuidade temporal da série do indicador.
anos_completos <- expand_grid(
  NU_ANO_CENSO = unique(c(igc_2013_2023$NU_ANO_CENSO, 2020)),
  CO_IES = unique(igc_2013_2023$CO_IES)
)
igc_2013_2023 <- full_join(
  igc_2013_2023,
  anos_completos,
  by = c("NU_ANO_CENSO", "CO_IES")
) %>%
  arrange(CO_IES, NU_ANO_CENSO)
# Estratégia de imputação:
# (i) para 2020, quando ausente, adota-se a média dos anos adjacentes
# (2019 e 2021), preservando a coerência temporal intra-institucional;
# (ii) persistindo valores ausentes, aplica-se a média da própria IES;
# (iii) na presença de lacunas residuais, aplica-se a média do ano.
igc_2013_2023 <- igc_2013_2023 %>%
  group_by(CO_IES) %>%
  mutate(
    IGC_CONTINUO = ifelse(
      NU_ANO_CENSO == 2020 & is.na(IGC_CONTINUO),
      mean(
        c(
          IGC_CONTINUO[NU_ANO_CENSO == 2019],
          IGC_CONTINUO[NU_ANO_CENSO == 2021]
        )
      ),
      na.rm = TRUE
    ),
    IGC_CONTINUO
  )
) %>%
  ungroup() %>%
  group_by(CO_IES) %>%

```

```

mutate(
  IGC_CONTINUO = ifelse(
    is.na(IGC_CONTINUO),
    mean(IGC_CONTINUO, na.rm = TRUE),
    IGC_CONTINUO
  )
) %>%
ungroup() %>%
group_by(NU_ANO_CENSO) %>%
mutate(
  IGC_CONTINUO = ifelse(
    is.na(IGC_CONTINUO),
    mean(IGC_CONTINUO, na.rm = TRUE),
    IGC_CONTINUO
  )
) %>%
ungroup()

```

Integração do IGC ao painel principal de IES e padronização da
variável temporal para 'Ano', em consonância com o restante do script.

```

dados_selecionados_ies <- dados_selecionados_ies %>%
left_join(
  igc_2013_2023 %>% select(NU_ANO_CENSO, CO_IES, IGC_CONTINUO),
  by = c("NU_ANO_CENSO", "CO_IES")
) %>%
rename(Ano = NU_ANO_CENSO)

```

```

# =====
# 1.5. TRATAMENTO DA BASE DE CURSOS
# =====

```

Objetivo:
Preparar a base de cursos para posterior consolidação no nível
DMU (IES-ano), gerando medidas agregadas de matrículas,
concluintes, abandono, financiamento estudantil e inclusão social.

```

# -----
# (i) Recorte consistente com a base de IES
# -----

```

Mantêm-se as mesmas categorias administrativas adotadas na base
institucional, assegurando coerência entre os bancos de dados.
Em seguida, constrói-se a variável indicadora de rede institucional
e padroniza-se o identificador do município de oferta do curso.

```

dados_cursos_2013_2023 <- dados_cursos_2013_2023 %>%
filter(TP_CATEGORIA_ADMINISTRATIVA %in% c(1, 4, 5, 6, 8, 9)) %>%
mutate(
  TP_REDE = if_else(TP_CATEGORIA_ADMINISTRATIVA == 1, 1, 2)
) %>%
rename(CO_MUN_OFERTA = CO_MUNICIPIO)

```

```

# -----
# (ii) Seleção de variáveis-chave e construção de derivados
# -----

```

São preservadas as variáveis essenciais ao processo educacional
e à modelagem de eficiência, incluindo matrículas, concluintes,
indicadores associados ao FIES, financiamento estudantil,

```

# reserva de vagas e abandono.
#
# Além disso, constrói-se:
# - QT_MAT_EAD: proxy de matrículas em EAD;
# - QT_CONC_EAD: proxy de concluintes em EAD;
# - QT_MAT_SOCIAL: proxy de inclusão social, combinando
# financiamento e reserva de vagas.

dados_selecionados_cursos <- dados_cursos_2013_2023 %>%
select(
  NU_ANO_CENSO, CO_IES, CO_MUN_OFERTA, TP_REDE, TP_DIMENSAO,
  QT_MAT, QT_MAT_DIURNO, QT_MAT_NOTURNO, QT_MAT_FIES,
  QT_CONC, QT_CONC_DIURNO, QT_CONC_NOTURNO, QT_CONC_FIES,
  QT_MAT_FINANC, QT_MAT_RESERVA_VAGA, QT_ABANDONO
) %>%
rename(Ano = NU_ANO_CENSO) %>%
mutate(
  QT_MAT_EAD = QT_MAT - QT_MAT_DIURNO - QT_MAT_NOTURNO,
  QT_CONC_EAD = QT_CONC - QT_CONC_DIURNO - QT_CONC_NOTURNO,
  QT_MAT_SOCIAL = QT_MAT_FINANC + QT_MAT_RESERVA_VAGA
)

# -----
# (iii) Consolidação no nível DMU (IES-ano)
# -----

# A base originalmente observada no nível de curso é agregada para o
# nível institucional por ano, compatibilizando-a com a estrutura
# analítica requerida para DEA/NDEA e análises ambientais subsequentes.

dados_totalizados_cursos <- dados_selecionados_cursos %>%
group_by(Ano, CO_IES) %>%
summarise(
  across(
    c(
      QT_MAT, QT_MAT_EAD, QT_MAT_FIES, QT_CONC, QT_CONC_EAD,
      QT_CONC_FIES, QT_MAT_FINANC, QT_MAT_SOCIAL, QT_ABANDONO
    ),
    \(x) sum(x, na.rm = TRUE)
  ),
  .groups = "drop"
)

# =====
# 1.6. JUNÇÃO FINAL DAS BASES E AJUSTES ESTRUTURAIIS
# =====

# Objetivo:
# Consolidar o painel final no nível IES-ano, integrando
# características institucionais, variáveis acadêmicas agregadas,
# indicador de qualidade e informação locacional.

# Função utilitária para geração de sigla institucional quando ausente.
# O procedimento remove preposições e utiliza as iniciais das palavras
# remanescentes, preservando identificação textual da IES em saídas
# analíticas, tabelas e figuras.

criar_sigla <- function(nome) {
  preps <- c("de", "da", "do", "das", "dos", "para", "em", "com", "por", "a", "e")

```

```

nome %>%
  str_split(" ") %>%
  unlist() %>%
  tolower() %>%
  .[!(. %in% preps)] %>%
  str_sub(1, 1) %>%
  paste(collapse = "") %>%
  toupper()
}

dados_sel_final <- dados_selecionados_ies %>%
  inner_join(dados_totalizados_cursos, by = c("Ano", "CO_IES")) %>%
  mutate(
    SG_IES = ifelse(
      is.na(SG_IES) | SG_IES == "",
      sapply(NO_IES, criar_sigla),
      SG_IES
    ),
    FLAG_IGC_NA = ifelse(is.na(IGC_CONTINUO), 1, 0)
  ) %>%
  select(
    Ano, CO_IES, NO_IES, SG_IES, TP_ORGANIZACAO_ACADEMICA, TP_REDE,
    TP_CATEGORIA_ADMINISTRATIVA, QT_DOC_EXE, QT_TEC_TOTAL,
    SOMATORIO_SERVICOS, SOMATORIO_ACERVO, QT_MAT, QT_MAT_EAD,
    QT_MAT_FIES, QT_CONC, QT_CONC_EAD, QT_CONC_FIES, QT_ABANDONO,
    QT_MAT_FINANC, QT_MAT_SOCIAL, IGC_CONTINUO, FLAG_IGC_NA,
    CO_MUNICIPIO_IES, NO_MUNICIPIO_IES, CO_MESORREGIAO_IES,
    NO_MESORREGIAO_IES, centroide
  )

# =====
# 1.7. IMPUTAÇÃO FINAL DO IGC POR TENDÊNCIA TEMPORAL
# =====

# Objetivo:
# Assegurar completude do indicador qualitativo no painel IES-ano,
# evitando perda de observações em etapas posteriores de modelagem.
#
# Estratégia de imputação:
# (i) interpolação linear intra-institucional (zoo::na.approx),
# preservando a trajetória temporal do IGC para cada IES;
# (ii) preenchimento residual por média anual, quando persistirem
# lacunas após a interpolação.

dados_sel_final <- dados_sel_final %>%
  arrange(CO_IES, Ano) %>%
  group_by(CO_IES) %>%
  mutate(
    IGC_CONTINUO = zoo::na.approx(
      IGC_CONTINUO,
      Ano,
      na.rm = FALSE,
      rule = 2
    )
  ) %>%
  ungroup() %>%
  group_by(Ano) %>%
  mutate(
    IGC_CONTINUO = ifelse(
      is.na(IGC_CONTINUO),

```

```

    mean(IGC_CONTINUO, na.rm = TRUE),
    IGC_CONTINUO
  )
) %>%
ungroup()

# =====
# 1.8. VALIDAÇÕES BÁSICAS DE INTEGRIDADE DO PAINEL
# =====

# Objetivo:
# Realizar checagens preliminares de consistência estrutural da base
# final, assegurando adequação do painel para as etapas subsequentes
# de modelagem, análise de eficiência e investigação de variáveis
# ambientais.

dim(dados_sel_final)          # dimensão final da base
glimpse(dados_sel_final)      # estrutura e tipos das variáveis
summary(dados_sel_final$IGC_CONTINUO) # distribuição do IGC contínuo
summary(dados_sel_final$QT_MAT)   # distribuição das matrículas
summary(dados_sel_final$QT_CONC)  # distribuição dos concluintes
range(dados_sel_final$Ano)       # intervalo temporal da base
length(unique(dados_sel_final$CO_IES)) # número de IES distintas
any(duplicated(dados_sel_final %>% select(Ano, CO_IES))) # unicidade IES-ano

# =====
# 1.9. FILTRO DOS ÚLTIMOS CINCO ANOS E PAINEL BALANCEADO
# =====

# Objetivo:
# Restringir o painel ao recorte temporal principal do estudo e,
# adicionalmente, identificar o subconjunto balanceado de IES
# presentes em todos os anos do período selecionado.

dados_sel_final_ult5 <- dados_sel_final %>%
  filter(Ano >= max(Ano) - 4)

# Frequência de observações por ano no recorte dos cinco anos mais recentes
dplyr::count(dados_sel_final_ult5, Ano)

# Identificação de IES com presença incompleta no recorte temporal
ies_presenca <- dados_sel_final_ult5 %>%
  count(CO_IES) %>%
  filter(n < length(unique(dados_sel_final_ult5$Ano)))

ies_presenca

# Construção do painel balanceado: somente IES presentes em todos os anos
ies_completas <- dados_sel_final_ult5 %>%
  group_by(CO_IES) %>%
  filter(n_distinct(Ano) == length(unique(dados_sel_final_ult5$Ano))) %>%
  ungroup()

count(ies_completas, Ano)

# =====
# 1.10. EXPORTAÇÃO DAS BASES TRATADAS
# =====

# Objetivo:

```

```
# Exportar as bases finais tratadas e consolidadas diretamente no
# diretório de trabalho principal, evitando dispersão de arquivos em
# múltiplas subpastas e preservando centralização do ambiente analítico.
```

```
out_dir <- base_dir
```

```
# -----
# Exportação em formato CSV
# -----
```

```
write_csv(
  dados_sel_final,
  file.path(out_dir, "dados_sel_final.csv")
)
```

```
write_csv(
  dados_sel_final_ult5,
  file.path(out_dir, "dados_sel_final_ult5.csv")
)
```

```
write_csv(
  ies_completas,
  file.path(out_dir, "dados_sel_final_ult5_balançado.csv")
)
```

```
# -----
# Exportação em formato XLSX
# -----
```

```
write_xlsx(
  dados_sel_final,
  path = file.path(out_dir, "dados_sel_final.xlsx")
)
```

```
write_xlsx(
  dados_sel_final_ult5,
  path = file.path(out_dir, "dados_sel_final_ult5.xlsx")
)
```

```
write_xlsx(
  ies_completas,
  path = file.path(out_dir, "dados_sel_final_ult5_balançado.xlsx")
)
```

```
# =====
# 1.11. PREPARAÇÃO DAS VARIÁVEIS PARA MODELAGEM DEA/NDEA
# =====
```

```
# Objetivo:
```

```
# Realizar o tratamento sistemático das variáveis quantitativas
# utilizadas na modelagem de eficiência. O procedimento envolve
# diagnóstico inicial, correção de valores ausentes ou nulos,
# transformações de escala, normalização e tratamento de outliers,
# assegurando maior estabilidade numérica e comparabilidade entre
# DMUs ao longo do painel temporal.
# Base inicial utilizada no tratamento
```

```
df <- dados_sel_final
```

```
# =====
# 1.11.1. DEFINIÇÃO DAS VARIÁVEIS DO NÚCLEO ANALÍTICO
# =====
```

```
# Seleciona o conjunto de variáveis quantitativas relevantes para
```

o modelo DEA/NDEA e para análises de sensibilidade posteriores.

```
vars_tratadas <- c(
  "QT_DOC_EXE",
  "QT_TEC_TOTAL",
  "QT_MAT",
  "QT_MAT_EAD",
  "QT_MAT_FIES",
  "QT_CONC",
  "QT_CONC_EAD",
  "QT_CONC_FIES",
  "QT_ABANDONO",
  "QT_MAT_FINANC",
  "QT_MAT_SOCIAL",
  "IGC_CONTINUO",
  "SOMATORIO_SERVICOS",
  "SOMATORIO_ACERVO"
)
```

Garante que apenas variáveis existentes na base sejam utilizadas
vars_tratadas <- vars_tratadas[vars_tratadas %in% names(df)]

Constante epsilon utilizada para evitar zeros e NA em etapas
posteriores de transformação
eps <- 1e-3

```
# =====
# 1.11.2. DIAGNÓSTICO INICIAL DE VALORES AUSENTES E ZEROS
# =====
```

Identifica a presença de valores ausentes (NA) e valores iguais a
zero nas variáveis selecionadas.

```
check_inicial <- tibble(
  Variavel = vars_tratadas,
  NAs = sapply(df[vars_tratadas], \(x) sum(is.na(x))),
  Zeros = sapply(df[vars_tratadas], \(x) sum(x == 0, na.rm = TRUE))
)
```

```
print(check_inicial)
```

```
# =====
# 1.11.3. AJUSTE DE NA E ZEROS VIA EPSILON
# =====
```

Substitui valores ausentes ou iguais a zero por epsilon,
garantindo positividade das variáveis utilizadas na modelagem.

```
df_trat <- df %>%
  mutate(
    across(
      all_of(vars_tratadas),
      ~ ifelse(is.na(.) | . == 0, eps, .),
      .names = "{.col}_trat"
    )
  )
```

```
vars_trat <- paste0(vars_tratadas, "_trat")
```

```

check_pos_eps <- tibble(
  Variavel = vars_trat,
  NAs = sapply(df_trat[vars_trat], \(x) sum(is.na(x))),
  Zeros = sapply(df_trat[vars_trat], \(x) sum(x == 0, na.rm = TRUE))
)

```

```
print(check_pos_eps)
```

```

# =====
# 1.11.4. TRANSFORMAÇÃO LOGARÍTMICA
# =====
# Aplica transformação logarítmica log10(x) para reduzir assimetrias
# distributivas e compressão de valores extremos.
df_trat <- df_trat %>%
  mutate(
    across(
      all_of(vars_trat),
      ~ log10(.),
      .names = "{.col}_log"
    )
  )
vars_trat_log <- paste0(vars_trat, "_log")

```

```

# =====
# 1.11.5. NORMALIZAÇÃO MIN-MAX NO INTERVALO [0,1] POR ANO
# =====
# Objetivo metodológico:
# Nesta etapa aplica-se uma transformação de normalização do tipo
# Min-Max às variáveis previamente tratadas e transformadas
# logaritmicamente, de modo a reescalá-las para o intervalo fechado
# [0,1]. Esse procedimento tem como finalidade reduzir a influência
# de diferenças de escala entre indicadores institucionais e
# garantir maior comparabilidade entre as DMUs (Instituições de
# Ensino Superior) no processo de estimação da eficiência.
#
# Fundamentação teórica:
# Embora os modelos DEA sejam, em princípio, invariantes à escala
# das variáveis, a normalização é frequentemente empregada em
# aplicações empíricas para mitigar problemas associados à presença
# de valores extremos, heterogeneidade de magnitude entre inputs e
# outputs e possíveis instabilidades numéricas decorrentes de
# transformações anteriores (como a transformação logarítmica).
#
# No contexto desta dissertação, a normalização Min-Max é aplicada
# separadamente para cada ano do painel, preservando a estrutura
# relativa de comparação entre as instituições em cada período
# analisado. Essa abordagem evita que variações estruturais entre
# anos distintos distorçam a escala das variáveis utilizadas na
# modelagem Network-DEA.
#
# A transformação utilizada segue a forma clássica:
#
# 
$$x_{norm} = (x - \min(x)) / (\max(x) - \min(x))$$

#
# Quando todos os valores da variável são idênticos em determinado
# ano (max = min), o procedimento retorna uma constante positiva
# pequena (eps), garantindo estabilidade numérica nas etapas
# subsequentes da modelagem.

```

```

#
# Referencial metodológico associado:
# - Cooper, Seiford e Zhu (2011) – Data Envelopment Analysis
# - Tone e Tsutsui (2009) – Network DEA
# - Simar e Wilson (2007) – Robustez e estabilidade em DEA
#
# A normalização é aplicada apenas às variáveis já tratadas nas
# etapas anteriores (log-transformadas), e os resultados são
# armazenados com o sufixo "_norm" para permitir rastreabilidade
# das transformações ao longo do pipeline de preparação dos dados.
normalize01 <- function(x, eps = 1e-3) {

  mx <- max(x, na.rm = TRUE)
  mn <- min(x, na.rm = TRUE)

  if (isTRUE(mx == mn)) {
    return(rep(eps, length(x)))
  }

  (x - mn) / (mx - mn)

}
df_trat <- df_trat %>%
  group_by(Ano) %>%
  mutate(
    across(
      all_of(vars_trat_log),
      ~ normalize01(., eps = eps),
      .names = "{.col}_norm"
    )
  ) %>%
  ungroup()
vars_log_norm <- paste0(vars_trat_log, "_norm")

# =====
# 1.11.6. TRATAMENTO DE OUTLIERS VIA WINSORIZAÇÃO (IQR)
# =====
# Objetivo metodológico:
# Nesta etapa aplica-se um procedimento de tratamento de valores
# extremos (outliers) por meio da técnica de winsorização baseada
# no intervalo interquartil (Interquartile Range – IQR). O objetivo
# é reduzir a influência desproporcional de observações atípicas
# que possam distorcer a distribuição das variáveis utilizadas na
# modelagem de eficiência.
#
# Fundamentação teórica:
# Em análises empíricas envolvendo Data Envelopment Analysis (DEA),
# a presença de valores extremos pode afetar significativamente a
# construção da fronteira eficiente, uma vez que a metodologia é
# sensível a observações que definem os limites do conjunto
# produtivo observado. Assim, valores atípicos podem gerar fronteiras
# artificialmente deslocadas e comprometer a interpretação dos
# escores de eficiência relativa entre as DMUs.
#
# A winsorização é um procedimento robusto que consiste em limitar
# os valores extremos de uma distribuição aos limites definidos por
# um critério estatístico. Neste estudo, utiliza-se o critério
# clássico baseado no intervalo interquartil:
#
# IQR = Q3 – Q1

```

```

#
# sendo Q1 o primeiro quartil (25%) e Q3 o terceiro quartil (75%).
# Os limites de truncamento são definidos como:
#
# Limite inferior =  $Q1 - 1.5 \times IQR$ 
# Limite superior =  $Q3 + 1.5 \times IQR$ 
#
# Valores abaixo do limite inferior são substituídos por esse
# próprio limite, enquanto valores acima do limite superior são
# substituídos pelo limite superior. Esse procedimento preserva a
# estrutura geral da distribuição, ao mesmo tempo em que reduz o
# impacto de observações extremas.
#
# No contexto desta dissertação, a winsorização é aplicada às
# variáveis previamente normalizadas no intervalo [0,1], sendo os
# resultados armazenados com o sufixo "_winsor". Após essa etapa,
# valores ausentes ou iguais a zero são substituídos por uma pequena
# constante positiva (eps), garantindo estabilidade numérica nas
# etapas subsequentes de estimação dos modelos DEA, Network-DEA e
# dos procedimentos bootstrap.
#
# Referencial metodológico associado:
# - Cooper, Seiford e Zhu (2011) – Data Envelopment Analysis
# - Simar e Wilson (2007) – Robust inference in DEA
# - Tukey (1977) – Exploratory Data Analysis (critério IQR)
#
# Dessa forma, o tratamento de outliers contribui para aumentar a
# robustez dos resultados empíricos e reduzir potenciais distorções
# na estimação da fronteira de eficiência.
winsor_iqr <- function(x) {

  q1 <- quantile(x, 0.25, na.rm = TRUE)
  q3 <- quantile(x, 0.75, na.rm = TRUE)

  iqr <- q3 - q1

  li <- q1 - 1.5 * iqr
  ls <- q3 + 1.5 * iqr

  x[x < li] <- li
  x[x > ls] <- ls

  x

}
df_trat <- df_trat %>%
mutate(
  across(
    all_of(vars_log_norm),
    winsor_iqr,
    .names = "{.col}_winsor"
  )
) %>%
mutate(
  across(
    ends_with("_winsor"),
    ~ ifelse(is.na(.) | . == 0, eps, .)
  )
)
vars_log_norm_winsor <- paste0(vars_log_norm, "_winsor")

```

```

# =====
# 1.11.7. CONSTRUÇÃO DA BASE FINAL PARA DEA/NDEA
# =====
# Objetivo metodológico:
# Esta etapa tem por finalidade consolidar, em um único objeto
# analítico, o conjunto de variáveis que servirá de base para a
# estimação dos modelos DEA, Network-DEA (NDEA), bootstrap manual
# e demais análises complementares desenvolvidas na dissertação.
#
# Fundamentação teórica:
# Em estudos de eficiência relativa, especialmente quando se
# pretende comparar diferentes especificações do modelo ou realizar
# testes de robustez, é recomendável preservar não apenas os dados
# originais, mas também suas versões sucessivamente transformadas.
# Isso permite avaliar a sensibilidade dos resultados a escolhas de
# pré-processamento, como correção de zeros, transformação
# logarítmica, normalização e tratamento de outliers.
#
# No contexto desta dissertação, a base final para DEA/NDEA é
# construída de forma a manter:
# (i) as variáveis originais;
# (ii) as variáveis ajustadas por epsilon (_trat);
# (iii) as variáveis transformadas por log (_trat_log);
# (iv) as variáveis normalizadas (_norm);
# (v) as variáveis winsorizadas (_winsor).
#
# Essa estratégia é metodologicamente relevante por duas razões
# principais:
# 1. garante rastreabilidade completa do processo de preparação dos
# dados, permitindo reproduzir cada etapa do pipeline analítico;
# 2. viabiliza análises de sensibilidade e comparação entre
# especificações do modelo, o que é especialmente importante em
# aplicações empíricas de DEA/NDEA com bases heterogêneas e
# suscetíveis à presença de assimetrias e valores extremos.
#
# Além das variáveis do núcleo analítico, são mantidos identificadores
# essenciais da DMU (Ano, CO_IES, SG_IES e TP_REDE), permitindo:
# - identificação institucional das unidades analisadas;
# - estratificação por tipo de rede (pública/privada);
# - consolidação temporal do painel;
# - vinculação posterior com resultados, gráficos e tabelas.
#
# A construção da base final é, portanto, a etapa que conecta o
# pré-processamento dos dados à modelagem de eficiência propriamente
# dita, constituindo o banco de trabalho utilizado nas estimções
# estáticas, nas análises em rede, no bootstrap e nas extensões
# posteriores, como o Índice de Malmquist e os modelos de aprendizado
# de máquina.
#
# Referencial metodológico associado:
# - Cooper, Seiford e Zhu (2011) – Data Envelopment Analysis
# - Tone e Tsutsui (2009) – Network DEA
# - Simar e Wilson (2007) – Robust inference in DEA
#
# O objeto resultante, denominado bd_DEA, representa a base analítica
# central da dissertação e serve como insumo para as etapas empíricas
# subsequentes.

cols_core <- c(

```

```

vars_tratadas,
vars_trat,
vars_trat_log,
vars_log_norm,
vars_log_norm_winsor
)

cols_id <- c(
  "Ano",
  "CO_IES",
  "SG_IES",
  "TP_REDE"
)

cols_keep <- unique(c(cols_id, cols_core))
cols_keep <- cols_keep[cols_keep %in% names(df_trat)]

bd_DEA <- df_trat %>%
  select(all_of(cols_keep))

# =====
# 1.11.8. EXPORTAÇÃO DA BASE FINAL
# =====
write_csv(
  bd_DEA,
  file.path(out_dir, "bd_final_DEA.csv")
)

write_xlsx(
  bd_DEA,
  path = file.path(out_dir, "bd_final_DEA.xlsx")
)
# =====
# 2. NDEA COM CORREÇÃO VIA BOOTSTRAP
# =====
# Objetivo:
# Estimar a eficiência das Instituições de Ensino Superior (IES) em
# uma estrutura de rede (Network Data Envelopment Analysis – NDEA),
# incorporando procedimentos de correção de viés via bootstrap,
# conforme a estratégia metodológica adotada na dissertação para
# obtenção de escores mais robustos, comparáveis e menos sensíveis
# a peculiaridades amostrais.
#
# =====
# 2.1. CARREGAMENTO DOS PACOTES
# =====
library(readr)
library(dplyr)
library(tidyr)
library(tibble)
library(deaR)
library(purrr)
library(writexl)
library(ggplot2)
library(sf)
library(geobr)
library(stringr)

# =====
# 2.2. DEFINIÇÃO DO PAINEL DE DADOS

```

```

# =====
# Objetivo metodológico:
# Esta etapa define o recorte temporal e a estrutura do painel de
# dados que será utilizada nas estimações de eficiência desenvolvidas
# na dissertação, restringindo a análise aos cinco anos mais recentes
# disponíveis na base consolidada.
#
# Fundamentação teórica:
# A adoção de um painel recente e balanceado é especialmente
# relevante em estudos de eficiência dinâmica, pois assegura maior
# comparabilidade intertemporal entre as DMUs e reduz distorções
# decorrentes da entrada ou saída de instituições ao longo do
# período de análise. Em aplicações de DEA/NDEA, essa padronização
# é importante para que mudanças nos escores reflitam, com maior
# precisão, alterações no desempenho relativo das unidades, e não
# apenas mudanças na composição da amostra.
#
# No contexto desta dissertação, optou-se por restringir a análise
# aos cinco anos mais recentes da série, formando inicialmente uma
# base parcial (dados_sel_final_ult5_base). Em seguida, mantém-se
# apenas o conjunto de instituições que possui observações em todos
# os anos desse recorte, resultando em um painel balanceado
# (dados_sel_final_balanceado_ult5).
#
# Essa opção metodológica é coerente com os objetivos empíricos do
# estudo, uma vez que:
# (i) melhora a consistência das comparações anuais;
# (ii) assegura base adequada para a análise de produtividade e para
# os índices intertemporais, como o Índice de Malmquist;
# (iii) evita que ausências esporádicas de informações comprometam a
# interpretação dos resultados do DEA/NDEA e do bootstrap.
#
# O vetor 'anos' armazena os exercícios efetivamente presentes no
# painel balanceado, sendo posteriormente utilizado para controlar
# iterações anuais no script e para garantir coerência entre as
# diferentes rotinas de estimação.
#
# Referencial metodológico associado:
# - Cooper, Seiford e Zhu (2011) – Data Envelopment Analysis
# - Tone e Tsutsui (2009) – Network DEA
# - Färe et al. (1994) – produtividade e comparações intertemporais
#
# O diretório de saída é mantido igual ao diretório principal
# definido no início do script, assegurando centralização dos
# arquivos gerados ao longo do pipeline analítico.

dados_sel_final_ult5_base <- dados_sel_final %>%
  filter(Ano >= max(Ano) - 4)

anos_esperados <- sort(unique(dados_sel_final_ult5_base$Ano))

dados_sel_final_balanceado_ult5 <- dados_sel_final_ult5_base %>%
  group_by(CO_IES) %>%
  filter(n_distinct(Ano) == length(anos_esperados)) %>%
  ungroup()

anos <- sort(unique(dados_sel_final_balanceado_ult5$Ano))

out_dir <- base_dir

```

```

# =====
# 2.3. FUNÇÕES AUXILIARES
# =====
# Objetivo metodológico:
# Esta seção define um conjunto de funções auxiliares utilizadas
# ao longo do processo de preparação dos dados e da estimação dos
# modelos de eficiência. Essas rotinas encapsulam operações
# recorrentes de transformação e padronização das variáveis,
# contribuindo para tornar o script mais modular, transparente
# e reproduzível.
#
# Fundamentação teórica:
# Em aplicações empíricas de Data Envelopment Analysis (DEA) e de
# suas extensões em rede (Network DEA – NDEA), é comum realizar
# ajustes preliminares nas variáveis de entrada (inputs) e saída
# (outputs) a fim de garantir estabilidade numérica e coerência
# analítica entre as unidades avaliadas. Entre esses ajustes
# destacam-se: tratamento de valores nulos ou iguais a zero,
# transformações logarítmicas para reduzir assimetrias e a
# padronização de variáveis categóricas para fins de interpretação
# dos resultados.
#
# No contexto desta dissertação, as funções auxiliares a seguir
# cumprem três finalidades principais:
#
# (i) substituição de valores iguais a zero por uma pequena constante
# positiva (epsilon), evitando problemas computacionais em
# transformações logarítmicas e em procedimentos de otimização
# presentes nos modelos DEA/NDEA;
#
# (ii) aplicação de transformação logarítmica suavizada, utilizada
# para reduzir a assimetria de variáveis com distribuição
# altamente concentrada ou com caudas longas, característica
# comum em bases institucionais de educação superior;
#
# (iii) rotulagem das categorias institucionais da variável TP_REDE,
# convertendo códigos numéricos em rótulos interpretáveis
# (por exemplo, "Pública" e "Privada (FIES)"), o que facilita
# a leitura de tabelas, gráficos e estatísticas comparativas
# apresentadas na análise empírica.
#
# A utilização dessas funções auxiliares permite padronizar o
# tratamento das variáveis ao longo de todo o pipeline analítico,
# evitando repetições de código e garantindo que os mesmos
# procedimentos sejam aplicados de forma consistente em todas as
# etapas da modelagem.
#
# Referencial metodológico associado:
# - Cooper, Seiford e Zhu (2011) – Data Envelopment Analysis
# - Tone e Tsutsui (2009) – Network DEA
# - Simar e Wilson (2007) – Robust inference in DEA
#
# Dessa forma, as funções apresentadas a seguir constituem elementos
# de suporte ao processamento das bases de dados e à interpretação
# dos resultados de eficiência obtidos nas seções subsequentes.

substituir_zero <- function(x) {
  if (is.numeric(x)) x[x == 0] <- 0.001
  x
}

```

```

transformar_log <- function(x) {
  if (is.numeric(x)) log(x + 1) else x
}

rotulo_rede <- function(x) {
  dplyr::case_when(
    x == 1 ~ "Pública",
    x == 2 ~ "Privada (FIES)",
    TRUE ~ as.character(x)
  )
}

# =====
# 2.4. FUNÇÃO DE ESTIMAÇÃO DEA POR ETAPA
# =====
# Objetivo metodológico:
# Esta função operacionaliza a estimação dos escores de eficiência
# técnica para cada etapa do modelo Network-DEA adotado na
# dissertação, considerando separadamente cada ano do painel e
# preservando a lógica sequencial do processo produtivo educacional.
#
# Fundamentação teórica:
# A modelagem em rede (Network DEA) parte do pressuposto de que o
# desempenho institucional não deve ser avaliado como um processo
# único e indiferenciado, mas sim como uma estrutura composta por
# subprocessos interligados, nos quais os resultados de uma etapa
# podem atuar como insumos ou produtos intermediários da etapa
# seguinte. Essa abordagem é especialmente adequada ao contexto das
# Instituições de Ensino Superior, cujo funcionamento envolve uma
# cadeia de transformação de recursos em resultados acadêmicos e
# institucionais.
#
# No contexto desta dissertação, a função "rodar_dea" estima, para
# cada ano e para cada estágio da rede, os escores de eficiência
# segundo um modelo DEA com orientação a produtos (output-oriented)
# e retornos variáveis de escala (Variable Returns to Scale – VRS).
#
# A escolha da orientação a outputs decorre do objetivo analítico do
# estudo: avaliar em que medida as instituições conseguem maximizar
# resultados acadêmicos e institucionais a partir dos recursos de
# que dispõem, sem pressupor redução proporcional dos insumos.
#
# A escolha de retornos variáveis de escala (VRS) é consistente com
# a heterogeneidade observada entre as IES brasileiras, que diferem
# substancialmente em porte, estrutura, capacidade operacional e
# contexto institucional. Nessa configuração, evita-se impor a
# hipótese de proporcionalidade estrita entre insumos e produtos,
# permitindo uma estimação mais aderente à realidade empírica.
#
# Estrutura do modelo em rede utilizada:
#
# Etapa 1:
# Recursos institucionais e infraestrutura informacional
# (acervo, serviços, docentes e técnicos) → Matrículas
#
# Etapa 2:
# Matrículas → Concluintes e Abandono
#
# Etapa 3:

```

```

# Concluintes → Qualidade institucional (IGC) e dimensão social
# (matrículas sociais)
#
# A função executa, em sequência, os seguintes procedimentos:
#
# (i) rotula o grupo institucional da DMU, distinguindo instituições
# públicas e privadas/FIES para fins de interpretação posterior;
#
# (ii) define, conforme a etapa do modelo, os inputs, outputs e
# eventuais produtos indesejáveis (undesirable outputs), como
# é o caso do abandono na Etapa 2;
#
# (iii) prepara a base de dados no formato requerido pelo pacote
# deaR, selecionando apenas as variáveis relevantes para a
# etapa considerada;
#
# (iv) aplica ajustes preliminares às variáveis, incluindo a
# substituição de zeros por pequena constante positiva e a
# transformação logarítmica, com o objetivo de reduzir
# assimetrias e assegurar estabilidade numérica;
#
# (v) estima o modelo DEA por meio da função "model_basic" do pacote
# deaR, configurado com orientação a outputs e retornos variáveis
# de escala;
#
# (vi) retorna os escores de eficiência no formato tabular,
# incluindo tanto o escore DEA original quanto sua forma
# invertida (1/eficiência), denominada no script como
# "Eficiencia_Ajustada".
#
# Interpretação dos escores:
# No modelo orientado a outputs implementado pelo pacote deaR, os
# escores retornados podem ser interpretados como medidas de distância
# à fronteira eficiente. O valor 1 representa eficiência técnica.
# Para padronizar a leitura dos resultados ao longo da dissertação,
# também é calculada a forma invertida do escore, permitindo
# interpretações comparativas adicionais no intervalo (0,1], em que
# valores mais próximos de 1 indicam melhor desempenho relativo.
#
# Referencial metodológico associado:
# - Charnes, Cooper e Rhodes (1978) – modelo DEA original
# - Banker, Charnes e Cooper (1984) – retornos variáveis de escala
# - Tone e Tsutsui (2009) – Network DEA
# - Cooper, Seiford e Zhu (2011) – Data Envelopment Analysis
#
# Dessa forma, a função "rodar_dea" constitui o núcleo operacional
# da estimação por estágio do modelo Network-DEA, servindo de base
# para as análises anuais, a construção da eficiência global da rede
# e os procedimentos posteriores de bootstrap e comparação entre
# grupos institucionais.

rodar_dea <- function(base_ano, ano, etapa) {

  base_ano <- base_ano %>%
    mutate(Grupo = rotulo_rede(TP_REDE))

  if (etapa == 1) {
    inputs <- c("SOMATORIO_ACERVO", "SOMATORIO_SERVICOS", "QT_DOC_EXE",
"QT_TEC_TOTAL")
    outputs <- c("QT_MAT")
  }
}

```

```

ud_inputs <- NULL
ud_outputs <- NULL

} else if (etapa == 2) {
  inputs <- c("QT_MAT")
  outputs <- c("QT_CONC", "QT_ABANDONO")
  ud_inputs <- NULL
  ud_outputs <- "QT_ABANDONO"

} else if (etapa == 3) {
  inputs <- c("QT_CONC")
  outputs <- c("IGC_CONTINUO", "QT_MAT_SOCIAL")
  ud_inputs <- NULL
  ud_outputs <- NULL
}

variaveis <- unique(c(inputs, outputs, ud_inputs, ud_outputs))

dados_dea <- base_ano %>%
  transmute(
    DMU = CO_IES,
    SG_IES,
    NO_IES,
    Grupo,
    Ano,
    across(all_of(variaveis), substituir_zero)
  ) %>%
  mutate(
    across(all_of(variaveis), transformar_log)
  )

args_make <- list(
  datadea = dados_dea,
  dmus = 1,
  inputs = inputs,
  outputs = outputs
)

if (!is.null(ud_inputs)) {
  args_make$ud_inputs <- match(ud_inputs, inputs)
}

if (!is.null(ud_outputs)) {
  args_make$ud_outputs <- match(ud_outputs, outputs)
}

dados_obj <- do.call(deaR::make_deadata, args_make)

modelo <- deaR::model_basic(
  dados_obj,
  orientation = "oo",
  rts = "vrs"
)

escore <- deaR::eficiencias(modelo)

tibble(
  DMU = dados_dea$DMU,
  SG_IES = dados_dea$SG_IES,
  NO_IES = dados_dea$NO_IES,

```

```

Grupo = dados_dea$Grupo,
Ano = ano,
Etapa = etapa,
Eficiencia = escore,
Eficiencia_Ajustada = 1 / escore
)
}

# =====
# 2.5. FUNÇÃO DE ANÁLISE DOS RESULTADOS POR ANO
# =====
# Objetivo metodológico:
# Esta função tem por finalidade sistematizar a análise dos resultados
# anuais obtidos na estimação DEA por etapa, sintetizando os escores
# de eficiência em estatísticas descritivas, rankings de desempenho e
# testes não paramétricos de comparação entre grupos institucionais.
#
# Fundamentação teórica:
# Em estudos empíricos baseados em DEA e Network-DEA, a obtenção dos
# escores de eficiência constitui apenas a primeira etapa da análise.
# Para que esses resultados possam ser interpretados de forma
# substantiva, é necessário organizá-los em medidas-resumo e
# instrumentos de comparação que permitam avaliar o comportamento das
# DMUs em cada período, identificar instituições de referência e
# verificar diferenças entre grupos.
#
# No contexto desta dissertação, a função "analisar_ano" cumpre esse
# papel ao processar, separadamente para cada ano e para cada etapa do
# modelo em rede, os resultados gerados pela função de estimação DEA.
# O procedimento contempla quatro dimensões analíticas principais:
#
# (i) estatísticas descritivas anuais:
# são calculadas medidas centrais e de dispersão dos escores de
# eficiência ajustada, incluindo média, mediana, valor mínimo e
# valor máximo. Essas estatísticas oferecem uma visão sintética do
# desempenho agregado das instituições em cada período analisado;
#
# (ii) rankings de desempenho:
# a função identifica as 10 instituições com melhor desempenho
# relativo (top 10) e as 10 com pior desempenho relativo
# (bottom 10), com base no escore de eficiência ajustada.
# Esse recurso é útil para fins de benchmarking, identificação de
# casos de referência e análise de heterogeneidade entre DMUs;
#
# (iii) testes não paramétricos de comparação entre grupos:
# são aplicados dois testes estatísticos para verificar se as
# distribuições dos escores diferem entre instituições públicas
# e privadas/FIES:
#
# - teste de Kolmogorov–Smirnov (KS):
# avalia se as distribuições empíricas dos grupos diferem entre
# si, sem pressupor normalidade;
#
# - teste de Kruskal–Wallis (KW):
# avalia diferenças de posição central entre grupos, sendo
# adequado para dados assimétricos ou distribuições não
# gaussianas.
#
# A escolha desses testes é consistente com a natureza dos
# escores DEA, que usualmente apresentam distribuição não normal,

```

```

#   assimetria e concentração em regiões próximas à fronteira;
#
# (iv) exportação dos resultados:
#   para cada combinação entre etapa e ano, a função exporta uma
#   planilha contendo os resultados completos, as estatísticas
#   descritivas e os rankings, o que favorece a transparência,
#   auditoria metodológica e reprodutibilidade do estudo.
#
# Interpretação dos escores:
# Neste script, a análise é conduzida com base na variável
# "Eficiencia_Ajustada", definida como o inverso do escore DEA. Essa
# padronização foi adotada para harmonizar a interpretação ao longo da
# dissertação e facilitar comparações entre etapas e entre diferentes
# abordagens empíricas. Assim, os rankings e estatísticas aqui
# apresentados refletem essa forma padronizada de mensuração.
#
# Referencial metodológico associado:
# - Cooper, Seiford e Zhu (2011) – Data Envelopment Analysis
# - Tone e Tsutsui (2009) – Network DEA
# - Siegel e Castellan (1988) – testes não paramétricos
#
# Dessa forma, a função "analisar_ano" constitui o elo entre a
# estimação dos escores de eficiência e sua interpretação empírica,
# permitindo transformar resultados brutos do DEA em evidências
# analíticas organizadas para comparação temporal e institucional.

analisar_ano <- function(resultado_ano, ano_alvo, etapa, out_dir) {

  stats_ano <- resultado_ano %>%
    summarise(
      Ano = ano_alvo,
      n_IES = n(),
      media = mean(Eficiencia_Ajustada, na.rm = TRUE),
      mediana = median(Eficiencia_Ajustada, na.rm = TRUE),
      min = min(Eficiencia_Ajustada, na.rm = TRUE),
      max = max(Eficiencia_Ajustada, na.rm = TRUE),
      .groups = "drop"
    )

  top10 <- resultado_ano %>%
    arrange(Eficiencia_Ajustada) %>%
    slice_head(n = 10)

  bottom10 <- resultado_ano %>%
    arrange(desc(Eficiencia_Ajustada)) %>%
    slice_head(n = 10)

  ks_pval <- tryCatch({
    if (length(unique(resultado_ano$Grupo)) > 1) {
      grupo_publica <- resultado_ano %>%
        filter(Grupo == "Pública") %>%
        pull(Eficiencia_Ajustada)

      grupo_privada <- resultado_ano %>%
        filter(Grupo == "Privada (FIES)") %>%
        pull(Eficiencia_Ajustada)

      if (length(grupo_publica) > 0 && length(grupo_privada) > 0) {
        ks.test(grupo_publica, grupo_privada)$p.value
      } else {

```

```

    NA_real_
  }
} else {
  NA_real_
}
}, error = function(e) NA_real_)

kw_pval <- tryCatch(
  kruskal.test(Eficiencia_Ajustada ~ Grupo, data = resultado_ano)$p.value,
  error = function(e) NA_real_
)

write_xlsx(
  list(
    resultados = resultado_ano,
    estatisticas = stats_ano,
    top10 = top10,
    bottom10 = bottom10
  ),
  path = file.path(out_dir, paste0("DEA_Etapa", etapa, "_", ano_alvo, ".xlsx"))
)

list(
  ano = ano_alvo,
  etapa = etapa,
  resultados = resultado_ano,
  stats = stats_ano,
  top10 = top10,
  bottom10 = bottom10,
  ks_pval = ks_pval,
  kw_pval = kw_pval
)
}

# =====
# 2.6. EXECUÇÃO DAS ESTIMAÇÕES POR ETAPA
# =====
# Objetivo metodológico:
# Esta seção executa a estimação dos modelos DEA para cada etapa da
# estrutura Network-DEA, considerando separadamente todos os anos do
# painel balanceado.
#
# Procedimento:
# Para cada etapa do modelo (1, 2 e 3), a função percorre o vetor de
# anos disponível no painel e aplica a função de estimação DEA
# definida anteriormente. Em seguida, os resultados anuais são
# processados pela função de análise, que gera estatísticas
# descritivas, rankings e arquivos de saída.
#
# Resultado:
# O procedimento retorna três objetos principais — um para cada etapa
# da rede — contendo:
# (i) resultados anuais completos,
# (ii) estatísticas agregadas, e
# (iii) base consolidada de escores de eficiência.
#
# Esses objetos são posteriormente utilizados para a construção da
# eficiência global da rede e para as análises empíricas
# subsequentes.

```

```

rodar_etapa_completa <- function(etapa) {

  resultados <- map(anos, function(a) {

    resultado_ano <- rodar_dea(
      dados_sel_final_balançado_ult5 %>% filter(Ano == a),
      ano = a,
      etapa = etapa
    )

    analisar_ano(resultado_ano, a, etapa, out_dir)
  })

  list(
    resultados_ano = resultados,
    stats_geral = map_dfr(resultados, "stats"),
    resultados_detalhados = map_dfr(resultados, "resultados")
  )
}

res_etapa1 <- rodar_etapa_completa(1)
res_etapa2 <- rodar_etapa_completa(2)
res_etapa3 <- rodar_etapa_completa(3)

# =====
# 2.7. CÁLCULO DA EFICIÊNCIA GLOBAL DA REDE
# =====
# Objetivo metodológico:
# Esta etapa calcula a eficiência global do modelo Network-DEA a
# partir da combinação dos escores de eficiência obtidos em cada uma
# das três etapas da rede produtiva.
#
# Procedimento:
# A eficiência global é construída pela multiplicação dos escores
# ajustados de cada estágio (Efic1, Efic2 e Efic3), refletindo o
# desempenho conjunto da instituição ao longo de toda a cadeia de
# transformação educacional.
#
# Resultado:
# O objeto resultante (eff_total) consolida os escores das três
# etapas e gera um indicador agregado de eficiência institucional
# para cada DMU e ano do painel. Em seguida, esses resultados são
# submetidos à mesma rotina de análise anual utilizada nas etapas
# individuais, permitindo estatísticas descritivas, rankings e
# exportação dos resultados da eficiência global da rede.

eff_total <- res_etapa1$resultados_detalhados %>%
  select(DMU, SG_IES, NO_IES, Grupo, Ano, Efic1 = Eficiencia_Ajustada) %>%
  left_join(
    res_etapa2$resultados_detalhados %>%
      select(DMU, Ano, Efic2 = Eficiencia_Ajustada),
    by = c("DMU", "Ano")
  ) %>%
  left_join(
    res_etapa3$resultados_detalhados %>%
      select(DMU, Ano, Efic3 = Eficiencia_Ajustada),
    by = c("DMU", "Ano")
  ) %>%
  mutate(
    Efic_Total_Ajustada = Efic1 * Efic2 * Efic3
  )

```

```

) %>%
left_join(
  dados_sel_final_balançado_ult5 %>%
  transmute(DMU = CO_IES, Ano, Grupo = rotulo_rede(TP_REDE)),
  by = c("DMU", "Ano")
)

map(anos, function(a) {

  resultado_ano <- eff_total %>%
  filter(Ano == a) %>%
  mutate(
    Eficiencia_Ajustada = Efic_Total_Ajustada
  )

  analisar_ano(resultado_ano, a, "Total", out_dir)
})

```

```

# =====
# 2.8. CONSOLIDAÇÃO DOS RESULTADOS PARA VISUALIZAÇÃO
# =====
# Objetivo metodológico:
# Esta etapa consolida, em uma única base analítica, os escores de
# eficiência obtidos nas três etapas do modelo Network-DEA e o
# indicador agregado de eficiência global da rede.
#
# Procedimento:
# Os resultados detalhados de cada etapa são combinados em formato
# longo (long format), adicionando a identificação da etapa
# correspondente. Em seguida, incorpora-se a eficiência global
# previamente calculada, garantindo estrutura uniforme para todas as
# observações.
#
# Resultado:
# O objeto resultante (eff_all) reúne, para cada DMU e ano do painel,
# os escores de eficiência ajustada por etapa e o indicador global
# da rede. Essa base consolidada é utilizada nas etapas posteriores
# de visualização gráfica, análises comparativas e construção das
# figuras e tabelas apresentadas na dissertação.

eff_all <- bind_rows(
  res_etapa1$resultados_detalhados %>% mutate(Etapa = "Etapa 1"),
  res_etapa2$resultados_detalhados %>% mutate(Etapa = "Etapa 2"),
  res_etapa3$resultados_detalhados %>% mutate(Etapa = "Etapa 3"),

  eff_total %>%
  left_join(
    dados_sel_final_balançado_ult5 %>%
    select(CO_IES, Ano, TP_REDE),
    by = c("DMU" = "CO_IES", "Ano" = "Ano")
  ) %>%
  mutate(
    Grupo = rotulo_rede(TP_REDE)
  ) %>%
  transmute(
    DMU,
    SG_IES,
    NO_IES,
    Grupo,
    Ano,

```

```

    Etapa = "Total",
    Eficiencia_Ajustada = Efic_Total_Ajustada
  )
)

# =====
# 2.9. GERAÇÃO DE GRÁFICOS DESCRITIVOS
# =====
# Objetivo metodológico:
# Esta etapa produz visualizações descritivas dos escores de eficiência
# obtidos pelo modelo Network-DEA, permitindo examinar a distribuição,
# dispersão e evolução temporal da eficiência das instituições.
#
# Procedimento:
# São gerados quatro tipos de gráficos com base na base consolidada
# de resultados (eff_all):
# (i) histograma da distribuição dos escores de eficiência ajustada,
# (ii) boxplot comparativo entre os grupos institucionais,
# (iii) série temporal da eficiência média agregada por ano e
# (iv) série temporal da eficiência média por grupo institucional.
#
# Resultado:
# Esses gráficos permitem avaliar padrões de distribuição dos escores,
# identificar diferenças estruturais entre instituições públicas e
# privadas e observar a evolução temporal do desempenho médio do
# sistema de ensino superior analisado.

g_hist_all <- ggplot(
  eff_all, aes(x = Eficiencia_Ajustada, fill = Grupo)) +
  geom_histogram(alpha = 0.7, bins = 30, color = "black") +
  labs(
    x = "Eficiência Ajustada (OO-VRS)",
    y = "Frequência"
  ) +
  theme_minimal()

g_box_group <- ggplot(
  eff_all, aes(x = Grupo, y = Eficiencia_Ajustada, fill = Grupo)) +
  geom_boxplot(alpha = 0.8) +
  theme_minimal()

media_por_ano <- eff_all %>%
  group_by(Ano) %>%
  summarise(
    media = mean(Eficiencia_Ajustada, na.rm = TRUE),
    .groups = "drop"
  )

g_media_ano <- ggplot(
  media_por_ano, aes(x = factor(Ano), y = media, group = 1)) +
  geom_line(linewidth = 1, color = "blue") +
  geom_point(size = 2, color = "red") +
  theme_minimal()

media_grupo_ano <- eff_all %>%
  group_by(Ano, Grupo) %>%
  summarise(
    media = mean(Eficiencia_Ajustada, na.rm = TRUE),
    .groups = "drop"
  )

g_media_grupo <- ggplot(
  media_grupo_ano, aes(x = factor(Ano), y = media, color = Grupo, group =
  Grupo)) +

```

```

geom_line(linewidth = 1) +
geom_point(size = 2) +
theme_minimal()

# =====
# 2.10. MAPA DE DISTRIBUIÇÃO ESPACIAL DAS IES (2019 E 2023)
# =====
# Objetivo metodológico:
# Esta etapa gera a visualização espacial das Instituições de Ensino
# Superior (IES) presentes na base de dados, permitindo observar a
# distribuição territorial das instituições públicas e privadas no
# Brasil.
#
# Procedimento:
# A partir da base consolidada de IES, são extraídas as coordenadas
# geográficas contidas no campo de centroide municipal. Em seguida,
# essas coordenadas são utilizadas para construir mapas temáticos
# sobre o território brasileiro utilizando o pacote geobr.
#
# Resultado:
# O gráfico final apresenta quatro mapas organizados em grade
# (2 anos × 2 redes administrativas), permitindo comparar a
# distribuição espacial das instituições públicas e privadas
# nos anos de 2019 e 2023.
#
# -----
# 2.10.1. Extração das coordenadas geográficas (longitude/latitude)
# -----
dados_mapa <- dados_sel_final_balanceado_ult5 %>%
  mutate(
    longitude = as.numeric(stringr::str_extract(centroide, "(?<=\\().*?(?=)")),
    latitude = as.numeric(stringr::str_extract(centroide, "(?<=).*?(?=\\)"))
  )

# -----
# 2.10.2. Obtenção do mapa base do Brasil
# -----
brasil <- geobr::read_country(year = 2020)

# -----
# 2.10.3. Preparação da base para os anos analisados
# -----
dados_plot <- dados_mapa %>%
  mutate(
    ano_plot = Ano,
    tp_rede_p = TP_REDE
  ) %>%
  filter(ano_plot %in% c(2019, 2023)) %>%
  mutate(
    rede_label = dplyr::case_when(
      tp_rede_p == 1 ~ "Públicas",
      tp_rede_p == 2 ~ "Privadas",
      TRUE ~ NA_character_
    ),
    rede_label = factor(rede_label, levels = c("Privadas", "Públicas")),
    ano_plot = factor(ano_plot, levels = c(2019, 2023))
  ) %>%
  filter(!is.na(rede_label))

# -----

```

```

# 2.10.4. Construção do mapa comparativo (4 painéis)
# -----
mapa_ies <- ggplot() +
  geom_sf(data = brasil, fill = "gray95", color = "gray70") +
  geom_point(
    data = dados_plot,
    aes(x = longitude, y = latitude, color = rede_label),
    size = 1.8,
    alpha = 0.75
  ) +
  facet_grid(ano_plot ~ rede_label) +
  scale_color_manual(values = c("Públicas" = "red", "Privadas" = "blue")) +
  labs(
    title = "Localização das Instituições de Ensino Superior (IES) no Brasil",
    subtitle = "Comparação entre 2019 e 2023 por rede administrativa",
    x = "Longitude",
    y = "Latitude"
  ) +
  coord_sf(expand = FALSE) +
  theme_minimal() +
  theme(
    legend.position = "none",
    panel.spacing = unit(1, "lines"),
    plot.title = element_text(face = "bold")
  )
)

mapa_ies
# =====
# 2.11. PACOTES COMPLEMENTARES PARA BOOTSTRAP E EXPORTAÇÃO
# =====
# Objetivo metodológico:
# Esta etapa assegura que o ambiente computacional contenha os
# pacotes necessários às rotinas complementares do estudo, em
# especial os procedimentos de bootstrap, exportação de resultados,
# modelagem adicional e análises auxiliares.
#
# Procedimento:
# O script verifica quais pacotes ainda não estão instalados no
# ambiente local e, quando necessário, realiza sua instalação antes
# do carregamento. Em seguida, todos os pacotes requeridos são
# carregados para a sessão de trabalho.
#
# Resultado:
# Essa rotina garante padronização do ambiente de execução,
# reprodutibilidade do pipeline analítico e disponibilidade das
# funções utilizadas nas etapas subsequentes da dissertação.
# Ao final, define-se uma semente aleatória fixa, assegurando
# replicabilidade dos procedimentos que envolvem reamostragem
# e particionamento dos dados.

pkgs <- c(
  "dplyr", "tidyr", "purrr", "readr", "tibble", "stringr",
  "ggplot2", "fmsb", "openxlsx", "writexl",
  "deaR", "Benchmarking", "rDEA", "pbapply",
  "janitor", "tidymodels", "recipes", "readxl"
)
to_install <- pkgs[!pkgs %in% rownames(installed.packages())]
if (length(to_install)) {
  install.packages(to_install, dependencies = TRUE)
}

```

```
invisible(lapply(pkgs, library, character.only = TRUE))
set.seed(123)

# =====
# 2.11.1. DEFINIÇÃO DE OBJETOS E FUNÇÕES DE APOIO
# =====
out_root <- base_dir
base_modelagem <- dados_sel_final_balanceado_ult5
anos <- sort(unique(base_modelagem$Ano))
dir_create <- function(...) {
  p <- file.path(...)
  dir.create(p, showWarnings = FALSE, recursive = TRUE)
  p
}
write_xlsx_list <- function(lst, path) {
  wb <- openxlsx::createWorkbook()

  for (nm in names(lst)) {
    openxlsx::addWorksheet(wb, nm)
    openxlsx::writeData(wb, nm, lst[[nm]])
  }

  openxlsx::saveWorkbook(wb, path, overwrite = TRUE)
}
# =====
# 2.11.2. ANÁLISE DESCRITIVA RÁPIDA
# =====
df <- bd_DEA
print(summary(df))
# =====
# 2.12. DEA ESTÁTICO POR EXERCÍCIO (dear) — ORIGINAL e TRATADAS
# =====
# Objetivo metodológico:
# Esta etapa estima os escores de eficiência por exercício,
# separadamente para cada estágio do processo produtivo modelado
# e para duas especificações dos dados: (i) variáveis originais e
# (ii) variáveis tratadas.
#
# Procedimento:
# São definidas duas estruturas de modelagem:
# - uma baseada nos dados originais;
# - outra baseada nas variáveis transformadas ao longo do pipeline
# de preparação dos dados.
#
# Para cada etapa, para cada hipótese de retornos de escala
# (CRS e VRS) e para cada ano do painel, o script estima um modelo
# DEA orientado a outputs utilizando o pacote dear. Além dos escores
# de eficiência, também são extraídos slacks, lambdas, targets e
# referências, que são exportados em planilhas individuais.
#
# Resultado:
# O procedimento gera, para cada combinação entre etapa, ano e tipo
# de retorno de escala, uma base de resultados detalhada e arquivos
# de saída que subsidiam a comparação entre modelos, a avaliação de
# robustez e a construção posterior dos indicadores globais de
# eficiência da rede.

out_dea_est <- dir_create(out_root, "03_DEA_Estatico_Resultados")
safe_df <- function(x) {
  if (is.null(x)) return(data.frame())
}
```

```

if (is.list(x)) {
  tryCatch(
    bind_rows(lapply(x, as.data.frame), .id = "parte"),
    error = function(e) as.data.frame(x)
  )
} else {
  as.data.frame(x)
}
}
DEA_SPEC_ORI <- list(
  etapa1 = list(
    inputs = c("QT_DOC_EXE", "QT_TEC_TOTAL"),
    outputs = c("QT_MAT"),
    ud_outputs = character(0)
  ),
  etapa2 = list(
    inputs = c("QT_MAT"),
    outputs = c("QT_CONC", "QT_ABANDONO"),
    ud_outputs = c("QT_ABANDONO")
  ),
  etapa3 = list(
    inputs = c("QT_CONC"),
    outputs = c("SOMATORIO_SERVICOS", "SOMATORIO_ACERVO"),
    ud_outputs = character(0)
  )
)

DEA_SPEC_TRA <- list(
  etapa1 = list(
    inputs = c("QT_DOC_EXE_trat_log_norm_winsor", "QT_TEC_TOTAL_trat_log_norm_winsor"),
    outputs = c("QT_MAT_trat_log_norm_winsor"),
    ud_outputs = character(0)
  ),
  etapa2 = list(
    inputs = c("QT_MAT_trat_log_norm_winsor"),
    outputs = c("QT_CONC_trat_log_norm_winsor", "QT_ABANDONO_trat_log_norm_winsor"),
    ud_outputs = c("QT_ABANDONO_trat_log_norm_winsor")
  ),
  etapa3 = list(
    inputs = c("QT_CONC_trat_log_norm_winsor"),
    outputs = c("SOMATORIO_SERVICOS_trat_log_norm_winsor",
"SOMATORIO_ACERVO_trat_log_norm_winsor"),
    ud_outputs = character(0)
  )
)

rodar_dea_ano <- function(df, ano, etapa_nome, inputs, outputs, ud_outputs, rts, out_prefix) {

  dat_y <- df %>%
    filter(Ano == ano) %>%
    select(DMU = CO_IES, all_of(inputs), all_of(outputs)) %>%
    tidyr::drop_na()

  if (nrow(dat_y) == 0) return(invisible(NULL))

  ud_idx <- which(outputs %in% ud_outputs)
  if (!length(ud_idx)) ud_idx <- NULL

  datadea <- deaR::make_deadata(
    dat_y,

```

```

    dmus = 1,
    inputs = 2:(1 + length(inputs)),
    outputs = (2 + length(inputs)):(1 + length(inputs) + length(outputs)),
    ud_outputs = ud_idx
  )

modelo <- deaR::model_basic(datadea, orientation = "oo", rts = rts)
eff_norm <- 1 / deaR::eficiencias(modelo)

sl <- deaR::slacks(modelo)
lam <- deaR::lambdas(modelo)
targ <- deaR::targets(modelo)
refs <- deaR::references(modelo)

wb <- openxlsx::createWorkbook()
openxlsx::addWorksheet(wb, "Scores")
openxlsx::writeData(wb, "Scores", data.frame(DMU = dat_y$DMU, Ef = eff_norm))

openxlsx::addWorksheet(wb, "Slacks")
openxlsx::writeData(wb, "Slacks", safe_df(sl))

openxlsx::addWorksheet(wb, "Lambdas")
openxlsx::writeData(wb, "Lambdas", safe_df(lam))

openxlsx::addWorksheet(wb, "Targets")
openxlsx::writeData(wb, "Targets", safe_df(targ))

openxlsx::addWorksheet(wb, "Refs")
openxlsx::writeData(wb, "Refs", safe_df(refs))

arq <- file.path(out_dea_est, sprintf("%s_%s_%s_OO_%s.xls", out_prefix, etapa_nome, rts, ano))
openxlsx::saveWorkbook(wb, arq, overwrite = TRUE)

message("✓ ", out_prefix, " | ", etapa_nome, " | ", rts, " | ano ", ano)

invisible(
  tibble(
    Ano = ano,
    DMU = dat_y$DMU,
    Etapa = etapa_nome,
    RTS = rts,
    Ef = eff_norm
  )
)
}

res_ori <- list()
for (etapa in names(DEA_SPEC_ORI)) {
  for (rts in c("crs", "vrs")) {
    for (a in anos) {
      tmp <- rodar_dea_ano(
        df, a, etapa,
        DEA_SPEC_ORI[[etapa]]$inputs,
        DEA_SPEC_ORI[[etapa]]$outputs,
        DEA_SPEC_ORI[[etapa]]$ud_outputs,
        rts,
        out_prefix = "Resultados_orig"
      )
      if (!is.null(tmp)) res_ori[[length(res_ori) + 1]] <- tmp
    }
  }
}

```

```

}
}

res_tra <- list()
for (etapa in names(DEA_SPEC_TRA)) {
  for (rts in c("crs", "vrs")) {
    for (a in anos) {
      tmp <- rodar_dea_ano(
        df, a, etapa,
        DEA_SPEC_TRA[[etapa]]$inputs,
        DEA_SPEC_TRA[[etapa]]$outputs,
        DEA_SPEC_TRA[[etapa]]$ud_outputs,
        rts,
        out_prefix = "Resultados_trat"
      )
      if (!is.null(tmp)) res_tra[[length(res_tra) + 1]] <- tmp
    }
  }
}

# =====
# 2.13. CONSOLIDAR (TRATADAS): EG e EE (CRS/VRS)
# =====
Data <- bind_rows(res_tra) %>%
tidyr::pivot_wider(
  id_cols = c(Ano, DMU, RTS),
  names_from = Etapa,
  values_from = Ef
) %>%
mutate(EG = etapa1 * etapa2 * etapa3)

EE_tbl <- Data %>%
select(Ano, DMU, RTS, EG) %>%
tidyr::pivot_wider(names_from = RTS, values_from = EG, names_prefix = "EG_") %>%
mutate(EE = EG_crs / EG_vrs)

Data_final <- Data %>%
left_join(EE_tbl %>% select(Ano, DMU, EE), by = c("Ano", "DMU"))

write_csv(Data_final, file.path(out_dea_est, "Eficiencias_Finais_tratadas_CRS_VRS.csv"))
write_xlsx_list(
  list("Eficiencias" = Data_final),
  file.path(out_dea_est, "Eficiencias_Finais_tratadas_CRS_VRS.xlsx")
)

# =====
# 2.14. OUTLIERS SIMPLES + EXPORTAÇÃO COMPLETA
# =====
# Objetivo metodológico:
# Esta etapa organiza e exporta os resultados consolidados do modelo
# NDEA, além de identificar observações com escores extremos de
# eficiência para fins de inspeção empírica.
#
# Procedimento:
# A base final de eficiências tratadas é recarregada a partir do
# arquivo consolidado, permitindo separar os escores por estágio e
# por hipótese de retornos de escala (CRS e VRS). Em seguida, são
# identificados casos extremos com base em um critério simples de
# corte, considerando escores muito baixos ou muito próximos da
# fronteira eficiente.

```

```

#
# Resultado:
# O procedimento gera:
# (i) planilhas consolidadas contendo os escores por estágio e por
# tipo de retorno de escala;
# (ii) listagens de observações extremas para cada etapa;
# (iii) gráficos boxplot para visualização sintética da dispersão dos
# escores sob VRS e CRS.

Data_reload <- read_csv(
  file.path(out_dea_est, "Eficiencias_Finais_tratadas_CRS_VRS.csv"),
  show_col_types = FALSE
)

vrs_e1 <- Data_reload %>% filter(RTS == "vrs") %>% select(Ano, DMU, RTS, Ef = etapa1)
vrs_e2 <- Data_reload %>% filter(RTS == "vrs") %>% select(Ano, DMU, RTS, Ef = etapa2)
vrs_e3 <- Data_reload %>% filter(RTS == "vrs") %>% select(Ano, DMU, RTS, Ef = etapa3)

outliers_e1 <- vrs_e1 %>% group_by(Ano) %>% filter(Ef < 0.2 | Ef > 0.99)
outliers_e2 <- vrs_e2 %>% group_by(Ano) %>% filter(Ef < 0.2 | Ef > 0.99)
outliers_e3 <- vrs_e3 %>% group_by(Ano) %>% filter(Ef < 0.2 | Ef > 0.99)

crs_e1 <- Data_reload %>% filter(RTS == "crs") %>% select(Ano, DMU, RTS, Ef = etapa1)
crs_e2 <- Data_reload %>% filter(RTS == "crs") %>% select(Ano, DMU, RTS, Ef = etapa2)
crs_e3 <- Data_reload %>% filter(RTS == "crs") %>% select(Ano, DMU, RTS, Ef = etapa3)

out_ndea <- dir_create(out_root, "04_NDEA_Resultados")

write_xlsx_list(
  list(
    "VRS_E1" = vrs_e1,
    "VRS_E2" = vrs_e2,
    "VRS_E3" = vrs_e3,
    "CRS_E1" = crs_e1,
    "CRS_E2" = crs_e2,
    "CRS_E3" = crs_e3,
    "EG_EE" = Data_reload %>% select(Ano, DMU, RTS, EG, EE),
    "Outliers_E1" = outliers_e1,
    "Outliers_E2" = outliers_e2,
    "Outliers_E3" = outliers_e3
  ),
  file.path(out_ndea, "NDEA_Eficiencias_CRS_VRS.xlsx")
)

png(file.path(out_ndea, "boxplot_estagios_vrs.png"), width = 1000, height = 300)
par(mfrow = c(1, 3), mar = c(5, 5, 2, 1))
boxplot(Data_reload$etapa1[Data_reload$RTS == "vrs"], main = "E1 (VRS)", ylab = "Eficiência")
boxplot(Data_reload$etapa2[Data_reload$RTS == "vrs"], main = "E2 (VRS)", ylab = "Eficiência")
boxplot(Data_reload$etapa3[Data_reload$RTS == "vrs"], main = "E3 (VRS)", ylab = "Eficiência")
dev.off()

png(file.path(out_ndea, "boxplot_estagios_crs.png"), width = 1000, height = 300)
par(mfrow = c(1, 3), mar = c(5, 5, 2, 1))
boxplot(Data_reload$etapa1[Data_reload$RTS == "crs"], main = "E1 (CRS)", ylab = "Eficiência")
boxplot(Data_reload$etapa2[Data_reload$RTS == "crs"], main = "E2 (CRS)", ylab = "Eficiência")
boxplot(Data_reload$etapa3[Data_reload$RTS == "crs"], main = "E3 (CRS)", ylab = "Eficiência")
dev.off()
# =====
# 2.15. ANÁLISE DESCRITIVA NDEA (grupos, gráficos)
# =====

```

```

out_desc <- dir_create(out_ndea, "4.1_Analise_Descritiva")
Data_reload <- Data_reload %>% mutate(DMU = as.character(DMU))
df <- df %>% mutate(CO_IES = as.character(CO_IES))
info_grupo <- df %>%
  group_by(CO_IES) %>%
  summarise(TP_REDE = first(TP_REDE), .groups = "drop") %>%
  mutate(Grupo = ifelse(TP_REDE == 1, "Pública", "Privada")) %>%
  select(DMU = CO_IES, Grupo)
Data_desc <- Data_reload %>%
  select(-matches("^Grupo")) %>%
  left_join(info_grupo, by = "DMU")
stats_por_grupo <- Data_desc %>%
  filter(is.finite(EG)) %>%
  group_by(Ano, Grupo, RTS) %>%
  summarise(
    N = n(),
    Minimo = min(EG),
    Mediana = median(EG),
    Media = mean(EG),
    Maximo = max(EG),
    Desvio_Padrao = sd(EG),
    .groups = "drop"
  )
g_hist_log <- ggplot(Data_desc, aes(x = EG)) +
  geom_histogram(bins = 30, fill = "darkgreen", alpha = .6, color = "black") +
  scale_x_log10() +
  labs(
    title = "Distribuição da Eficiência Global (log10)",
    x = "EG (log10)",
    y = "Frequência"
  ) +
  theme_minimal()
ggsave(file.path(out_desc, "NDEA_Histograma_Log10.png"), g_hist_log, width = 9, height = 6, dpi =
300)
EG_long <- Data_desc %>%
  select(Ano, Grupo, etapa1, etapa2, etapa3, EG) %>%
  pivot_longer(c(etapa1, etapa2, etapa3, EG), names_to = "Etapa", values_to = "Eficiencia") %>%
  mutate(
    Etapa = recode(
      Etapa,
      "etapa1" = "Etapa 1",
      "etapa2" = "Etapa 2",
      "etapa3" = "Etapa 3",
      "EG" = "Total"
    )
  )
g_hist_ge <- ggplot(EG_long, aes(x = Eficiencia, fill = Grupo)) +
  geom_histogram(bins = 30, alpha = .6, position = "identity", color = "black") +
  geom_vline(xintercept = 1, linetype = "dashed", color = "red", linewidth = 1) +
  scale_x_log10() +
  facet_wrap(~Etapa, ncol = 2, scales = "free_y") +
  labs(
    title = "Eficiências por Grupo e Etapa (log10)",
    x = "Eficiência (log10)",
    y = "Frequência"
  ) +
  theme_minimal() +
  scale_fill_manual(values = c("Pública" = "darkgreen", "Privada" = "darkblue"))
ggsave(file.path(out_desc, "NDEA_Hist_Grupo_Etapas_Log10.png"), g_hist_ge, width = 12, height =
8, dpi = 300)

```

```

stats_grupo_ano <- EG_long %>%
  filter(is.finite(Eficiencia)) %>%
  group_by(Ano, Grupo, Etapa) %>%
  summarise(
    Media = mean(Eficiencia),
    Mediana = median(Eficiencia),
    .groups = "drop"
  )
anos_unicos <- sort(unique(stats_grupo_ano$Ano))
g_linha <- ggplot(stats_grupo_ano, aes(x = Ano, color = Grupo, group = Grupo)) +
  geom_line(aes(y = Media, linetype = "Média"), linewidth = 1.2) +
  geom_point(aes(y = Media), size = 2) +
  geom_line(aes(y = Mediana, linetype = "Mediana"), linewidth = 1.2) +
  geom_point(aes(y = Mediana), shape = 17, size = 2) +
  facet_wrap(~Etapa, ncol = 2, scales = "free_y") +
  labs(
    title = "Evolução da Eficiência (Média x Mediana)",
    x = "Ano",
    y = "Eficiência"
  ) +
  scale_x_continuous(breaks = anos_unicos) +
  scale_color_manual(values = c("Pública" = "darkgreen", "Privada" = "darkblue")) +
  scale_linetype_manual(values = c("Média" = "solid", "Mediana" = "dashed")) +
  theme_minimal()
ggsave(file.path(out_desc, "NDEA_Linhas_Grupo_Ano_Etapas.png"), g_linha, width = 12, height = 8,
  dpi = 300)
g_box <- EG_long %>%
  ggplot(aes(x = factor(Ano), y = Eficiencia, fill = Grupo)) +
  geom_boxplot(outlier.alpha = .2, position = position_dodge(.8)) +
  facet_wrap(~Etapa, ncol = 2, scales = "free_y") +
  scale_y_log10() +
  labs(
    title = "Boxplot por Ano e Grupo",
    x = "Ano",
    y = "Eficiência (log10)"
  ) +
  scale_fill_manual(values = c("Pública" = "darkgreen", "Privada" = "darkblue")) +
  theme_minimal()
ggsave(file.path(out_desc, "NDEA_Boxplot_Ano_Grupo_Etapas.png"), g_box, width = 14, height = 8,
  dpi = 300)
writexl::write_xlsx(
  list(
    "Estatisticas_Por_Grupo" = stats_por_grupo,
    "NDEA_Resultados" = Data_desc,
    "Evolucao_Media_Mediana" = stats_grupo_ano
  ),
  path = file.path(out_desc, "NDEA_Analise_Descritiva.xlsx")
)
# =====
# 2.16. SUPEREFICIÊNCIA (VRS, 3 estágios) com CHECKPOINTS
# =====
# Objetivo metodológico:
# Esta etapa estima medidas de supereficiência para cada estágio do
# modelo NDEA sob retornos variáveis de escala (VRS), com o objetivo
# de refinar a discriminação entre DMUs classificadas como eficientes
# no modelo DEA tradicional.
#
# Procedimento:
# Para cada ano do painel e para cada um dos três estágios, o script
# estima inicialmente os escores DEA usuais e, em seguida, calcula os

```

```

# escores de supereficiência por meio da função sdea() do pacote
# Benchmarking. Os resultados anuais são salvos em arquivos
# intermediários (checkpoints), permitindo retomar a execução sem
# necessidade de reprocessar anos já concluídos.
#
# Resultado:
# O procedimento gera:
# (i) tabelas anuais e consolidadas de supereficiência por estágio;
# (ii) um indicador agregado de supereficiência global (SEG),
#     calculado pelo produto dos escores dos três estágios;
# (iii) estatísticas-resumo anuais para os casos com SEG superior a 1,
#     permitindo identificar instituições com desempenho acima da
#     fronteira eficiente convencional.
#
out_sup <- dir_create(out_ndea, "4_Supereficiência")
superef_all_e1 <- list()
superef_all_e2 <- list()
superef_all_e3 <- list()
for (a in anos) {
  cat("→ □ Supereficiência | Ano:", a, "\n")

  dfa <- df %>% filter(Ano == a)

  f1 <- file.path(out_sup, paste0("Superef_E1_", a, ".csv"))
  f2 <- file.path(out_sup, paste0("Superef_E2_", a, ".csv"))
  f3 <- file.path(out_sup, paste0("Superef_E3_", a, ".csv"))

  if (file.exists(f1) && file.exists(f2) && file.exists(f3)) {
    superef_all_e1[[as.character(a)]] <- read_csv(f1, show_col_types = FALSE)
    superef_all_e2[[as.character(a)]] <- read_csv(f2, show_col_types = FALSE)
    superef_all_e3[[as.character(a)]] <- read_csv(f3, show_col_types = FALSE)
    next
  }

  x1 <- as.matrix(dfa[, c("QT_DOC_EXE_trat_log_norm_winsor",
"QT_TEC_TOTAL_trat_log_norm_winsor")])
  y1 <- as.matrix(dfa[, "QT_MAT_trat_log_norm_winsor", drop = FALSE])
  dea_e1 <- Benchmarking::dea(x1, y1, RTS = "vrs", ORIENTATION = "out")
  se_e1 <- Benchmarking::sdea(x1, y1, RTS = "vrs", ORIENTATION = "out")
  dfse1 <- data.frame(CO_IES = dfa$CO_IES, NO_IES = dfa$SG_IES, Ano = a, DEA_E1 = 1 /
  dea_e1$eff, SE_E1 = 1 / se_e1$eff)

  x2 <- as.matrix(dfa[, c("QT_MAT_trat_log_norm_winsor",
"QT_ABANDONO_trat_log_norm_winsor")])
  y2 <- as.matrix(dfa[, "QT_CONC_trat_log_norm_winsor", drop = FALSE])
  dea_e2 <- Benchmarking::dea(x2, y2, RTS = "vrs", ORIENTATION = "out")
  se_e2 <- Benchmarking::sdea(x2, y2, RTS = "vrs", ORIENTATION = "out")
  dfse2 <- data.frame(CO_IES = dfa$CO_IES, NO_IES = dfa$SG_IES, Ano = a, DEA_E2 = 1 /
  dea_e2$eff, SE_E2 = 1 / se_e2$eff)

  x3 <- as.matrix(dfa[, "QT_CONC_trat_log_norm_winsor", drop = FALSE])
  y3 <- as.matrix(dfa[, c("SOMATORIO_SERVICOS_trat_log_norm_winsor",
"SOMATORIO_ACERVO_trat_log_norm_winsor")])
  dea_e3 <- Benchmarking::dea(x3, y3, RTS = "vrs", ORIENTATION = "out")
  se_e3 <- Benchmarking::sdea(x3, y3, RTS = "vrs", ORIENTATION = "out")
  dfse3 <- data.frame(CO_IES = dfa$CO_IES, NO_IES = dfa$SG_IES, Ano = a, DEA_E3 = 1 /
  dea_e3$eff, SE_E3 = 1 / se_e3$eff)

  superef_all_e1[[as.character(a)]] <- dfse1
  superef_all_e2[[as.character(a)]] <- dfse2

```

```

superef_all_e3[[as.character(a)]] <- dfse3

write_csv(dfse1, f1)
write_csv(dfse2, f2)
write_csv(dfse3, f3)

cat("✓ Checkpoints salvos:", a, "\n")
}
superef_final_e1 <- bind_rows(superef_all_e1)
superef_final_e2 <- bind_rows(superef_all_e2)
superef_final_e3 <- bind_rows(superef_all_e3)
write_xlsx_list(
  list(
    "Superef_E1" = superef_final_e1,
    "Superef_E2" = superef_final_e2,
    "Superef_E3" = superef_final_e3
  ),
  file.path(out_sup, "NDEA_Superef_VRS.xlsx")
)
seg_tabela <- superef_final_e1 %>%
  select(CO_IES, NO_IES, Ano, SE_E1) %>%
  inner_join(superef_final_e2 %>% select(CO_IES, Ano, SE_E2), by = c("CO_IES", "Ano")) %>%
  inner_join(superef_final_e3 %>% select(CO_IES, Ano, SE_E3), by = c("CO_IES", "Ano")) %>%
  mutate(SEG = SE_E1 * SE_E2 * SE_E3)
seg_filtrado <- seg_tabela %>% filter(SEG > 1)
seg_stats <- seg_filtrado %>%
  group_by(Ano) %>%
  summarise(
    N = n(),
    Media_SEG = mean(SEG),
    Mediana_SEG = median(SEG),
    Max_SEG = max(SEG),
    Min_SEG = min(SEG),
    .groups = "drop"
  )
out_seg <- dir_create(out_ndea, "4.1_Superef_Geral")
write_xlsx_list(
  list(
    "SEG_Detalhado" = seg_filtrado,
    "SEG_Resumo" = seg_stats
  ),
  file.path(out_seg, "SuperefGeral_SEG_maior1.xlsx")
)
# =====
# 2.17. TESTE DO MODELO (CRS vs VRS) — 3 etapas + KS
# =====
# Objetivo metodológico:
# Esta etapa compara as duas principais hipóteses de retornos de
# escala utilizadas em DEA — constantes (CRS) e variáveis (VRS) —
# com o objetivo de avaliar qual especificação é mais adequada para
# representar a tecnologia de produção das instituições analisadas.
#
# Procedimento:
# Para cada ano e para cada um dos três estágios do modelo, são
# estimados escores de eficiência sob CRS e VRS. A partir desses
# resultados, calculam-se:
# (i) a eficiência de escala;
# (ii) testes de Kolmogorov–Smirnov (KS) para comparar a distribuição
# dos escores CRS e VRS;
# (iii) uma tabela-resumo com estatísticas médias anuais e indicação

```

```

# do modelo mais aderente.
#
# Resultado:
# O procedimento produz bases consolidadas por etapa e uma tabela
# final de apoio à decisão metodológica, permitindo justificar a
# adoção do modelo VRS ou CRS no restante da dissertação.
#
out_test <- dir_create(out_ndea, "6_Teste_Modelo")
rodar_dea_modelos <- function(dfa, x_vars, y_vars) {
  x <- as.matrix(dfa[, x_vars, drop = FALSE])
  y <- as.matrix(dfa[, y_vars, drop = FALSE])

  crs <- Benchmarking::dea(x, y, RTS = "crs", ORIENTATION = "out")
  vrs <- Benchmarking::dea(x, y, RTS = "vrs", ORIENTATION = "out")

  Eff_CRS <- 1 / crs$eff
  Eff_VRS <- 1 / vrs$eff

  tibble(
    CO_IES = dfa$CO_IES,
    SG_IES = dfa$SG_IES,
    Ano = dfa$Ano,
    Eff_CRS = Eff_CRS,
    Eff_VRS = Eff_VRS,
    Eff_Scale = Eff_VRS / Eff_CRS
  )
}
teste_e1 <- list()
teste_e2 <- list()
teste_e3 <- list()
chk_dir <- dir_create(out_test, "checkpoints")
for (a in anos) {
  dfa <- df %>% filter(Ano == a)

  f1 <- file.path(chk_dir, paste0("teste_e1_", a, ".rds"))
  if (file.exists(f1)) {
    res_e1 <- readRDS(f1)
  } else {
    res_e1 <- rodar_dea_modelos(
      dfa,
      c("QT_DOC_EXE_trat_log_norm_winsor", "QT_TEC_TOTAL_trat_log_norm_winsor"),
      "QT_MAT_trat_log_norm_winsor"
    )
    saveRDS(res_e1, f1)
  }
  teste_e1[[as.character(a)]] <- res_e1

  f2 <- file.path(chk_dir, paste0("teste_e2_", a, ".rds"))
  if (file.exists(f2)) {
    res_e2 <- readRDS(f2)
  } else {
    res_e2 <- rodar_dea_modelos(
      dfa,
      c("QT_MAT_trat_log_norm_winsor", "QT_ABANDONO_trat_log_norm_winsor"),
      "QT_CONC_trat_log_norm_winsor"
    )
    saveRDS(res_e2, f2)
  }
  teste_e2[[as.character(a)]] <- res_e2
}

```

```

f3 <- file.path(chk_dir, paste0("teste_e3_", a, ".rds"))
if (file.exists(f3)) {
  res_e3 <- readRDS(f3)
} else {
  res_e3 <- rodar_dea_modelos(
    dfa,
    "QT_CONC_trat_log_norm_winsor",
    c("SOMATORIO_SERVICOS_trat_log_norm_winsor",
      "SOMATORIO_ACERVO_trat_log_norm_winsor")
  )
  saveRDS(res_e3, f3)
}
teste_e3[[as.character(a)]] <- res_e3
}
teste_final_e1 <- bind_rows(teste_e1)
teste_final_e2 <- bind_rows(teste_e2)
teste_final_e3 <- bind_rows(teste_e3)
ks_e1 <- ks.test(teste_final_e1$Eff_CRS, teste_final_e1$Eff_VRS)
ks_e2 <- ks.test(teste_final_e2$Eff_CRS, teste_final_e2$Eff_VRS)
ks_e3 <- ks.test(teste_final_e3$Eff_CRS, teste_final_e3$Eff_VRS)
print(ks_e1)
print(ks_e2)
print(ks_e3)
write_xlsx_list(
  list(
    "Teste_Etapa1" = teste_final_e1,
    "Teste_Etapa2" = teste_final_e2,
    "Teste_Etapa3" = teste_final_e3
  ),
  file.path(out_test, "NDEA_Testes_Modelos_CRSvsVRS.xlsx")
)
ks_tabela <- map_dfr(anos, function(a) {
  ks1 <- ks.test(teste_e1[[as.character(a)]]$Eff_CRS, teste_e1[[as.character(a)]]$Eff_VRS)
  ks2 <- ks.test(teste_e2[[as.character(a)]]$Eff_CRS, teste_e2[[as.character(a)]]$Eff_VRS)
  ks3 <- ks.test(teste_e3[[as.character(a)]]$Eff_CRS, teste_e3[[as.character(a)]]$Eff_VRS)

  ratio <- mean(teste_e1[[as.character(a)]]$Eff_VRS, na.rm = TRUE) /
    mean(teste_e1[[as.character(a)]]$Eff_CRS, na.rm = TRUE)

  modelo_adequado <- ifelse(
    all(c(ks1$p.value, ks2$p.value, ks3$p.value) < .05),
    "VRS",
    "CRS"
  )

  tibble(
    Ano = a,
    Ratio = ratio,
    KS_Estatistica = mean(c(ks1$statistic, ks2$statistic, ks3$statistic)),
    KS_p_valor = mean(c(ks1$p.value, ks2$p.value, ks3$p.value)),
    Modelo_Adequado = modelo_adequado
  )
})
write_xlsx_list(
  list("KS_Resultados" = ks_tabela),
  file.path(out_test, "Teste_KS_Resultados.xlsx")
)
# =====
# 2.18. KRUSKAL-WALLIS (EG, VRS) — por ano e global
# =====

```

```

# Objetivo metodológico:
# Esta etapa avalia se há diferenças estatisticamente significativas
# na eficiência global (EG), sob retornos variáveis de escala (VRS),
# entre os grupos institucionais considerados na pesquisa.
#
# Procedimento:
# Inicialmente, as DMUs são classificadas em grupos conforme o tipo
# de rede institucional. Em seguida, aplica-se o teste não
# paramétrico de Kruskal–Wallis para:
# (i) comparar os grupos em cada ano do painel e
# (ii) comparar os grupos no conjunto agregado de todos os anos.
#
# Resultado:
# O procedimento gera tabelas com os valores da estatística do teste,
# graus de liberdade, p-valores e indicação de significância
# estatística, permitindo verificar se a eficiência global difere
# entre os grupos institucionais ao longo do período analisado.
#
out_kw <- dir_create(out_nde, "8_Teste_Kruskal_Wallis")
df_mapa <- df %>%
  group_by(CO_IES) %>%
  summarise(
    TP_REDE = first(TP_REDE),
    .groups = "drop"
  ) %>%
  transmute(
    DMU = CO_IES,
    Grupo = case_when(
      TP_REDE == 1 ~ "Pública",
      TP_REDE == 2 ~ "Privada",
      TRUE ~ "Não Classificado"
    )
  )
EG_desc <- Data_reload %>%
  left_join(df_mapa, by = "DMU")
EG_vrs <- EG_desc %>%
  filter(RTS == "vrs")
kw_por_ano_df <- map_dfr(sort(unique(EG_vrs$Ano)), function(a) {
  dfa <- EG_vrs %>% filter(Ano == a)
  if (length(unique(dfa$Grupo)) > 1) {
    teste <- kruskal.test(EG ~ Grupo, data = dfa)
    tibble(
      Ano = a,
      Chi_Sq = as.numeric(teste$statistic),
      DF = as.numeric(teste$parameter),
      p_valor = teste$p.value,
      Significativo = ifelse(teste$p.value < .05, "Sim", "Não")
    )
  } else {
    tibble(Ano = a, Chi_Sq = NA, DF = NA, p_valor = NA, Significativo = "NA")
  }
})
kw_global <- kruskal.test(EG ~ Grupo, data = EG_vrs)
kw_global_df <- tibble(
  Ano = "Todos",
  Chi_Sq = as.numeric(kw_global$statistic),
  DF = as.numeric(kw_global$parameter),
  p_valor = kw_global$p.value,
  Significativo = ifelse(kw_global$p.value < .05, "Sim", "Não")
)

```

```

kw_final <- bind_rows(kw_por_ano_df %>% mutate(Ano = as.character(Ano)), kw_global_df)
write_xlsx_list(
  list(
    "Kruskal_Ano" = kw_por_ano_df,
    "Kruskal_Global" = kw_global_df,
    "Kruskal_Final" = kw_final
  ),
  file.path(out_kw, "Teste_Kruskal_Wallis_EG.xlsx")
)
# =====
# 2.19. BOOTSTRAP MANUAL (Simar & Wilson) — 3 estágios (VRS, out)
# =====
# Objetivo metodológico:
# Esta etapa implementa um procedimento de bootstrap manual para os
# três estágios do modelo NDEA, com o objetivo de corrigir o viés dos
# escores de eficiência estimados no modelo determinístico.
#
# Procedimento:
# Para cada ano e para cada estágio da rede, são realizadas
# reamostragens com reposição sobre a base observada, gerando novas
# fronteiras de referência sob retornos variáveis de escala (VRS) e
# orientação a outputs. A partir dessas reamostragens, calculam-se:
# (i) o escore observado,
# (ii) o escore corrigido por viés,
# (iii) os intervalos de confiança empíricos.
#
# Resultado:
# O procedimento gera bases anuais e consolidadas com os escores
# bootstrap por estágio, uma tabela-resumo das diferenças entre o
# modelo clássico e o modelo corrigido, e tabelas finais que
# permitem comparar diretamente os resultados do NDEA determinístico
# e do NDEA com correção bootstrap.
#
out_boot <- dir_create(out_ndea, "7_NDEA_Bootstrap_Manual")
B <- 500
bootstrap_dea_vrs_out <- function(X, Y, B = 500, conf = c(.025, .975)) {
  n <- nrow(X)
  base_fit <- Benchmarking::dea(X, Y, RTS = "vrs", ORIENTATION = "out")
  theta_hat <- as.numeric(base_fit$eff)
  theta_star <- matrix(NA_real_, nrow = n, ncol = B)
  idx_all <- seq_len(n)

  for (b in 1:B) {
    idx <- sample(idx_all, size = n, replace = TRUE)
    Xb <- X[idx, , drop = FALSE]
    Yb <- Y[idx, , drop = FALSE]
    fit_b <- Benchmarking::dea(X, Y, XREF = Xb, YREF = Yb, RTS = "vrs", ORIENTATION = "out")
    theta_star[, b] <- as.numeric(fit_b$eff)
  }

  theta_star_mean <- rowMeans(theta_star, na.rm = TRUE)
  theta_bc <- 2 * theta_hat - theta_star_mean
  theta_ci_low <- apply(theta_star, 1, quantile, conf[1], na.rm = TRUE)
  theta_ci_high <- apply(theta_star, 1, quantile, conf[2], na.rm = TRUE)

  tibble(
    Eff_Obs = 1 / theta_hat,
    Eff_BC = 1 / theta_bc,
    CI_Lower = 1 / theta_ci_high,
    CI_Upper = 1 / theta_ci_low,

```

```

    theta = theta_hat,
    theta_bc = theta_bc
  )
}
boot_e1_all <- list()
boot_e2_all <- list()
boot_e3_all <- list()
resumo_boot <- list()
for (a in anos) {
  cat("→ □ Bootstrap | Ano:", a, "\n")

  dfa <- df %>% filter(Ano == a)

  f1 <- file.path(out_boot, paste0("Boot_E1_", a, ".csv"))
  f2 <- file.path(out_boot, paste0("Boot_E2_", a, ".csv"))
  f3 <- file.path(out_boot, paste0("Boot_E3_", a, ".csv"))
  fR <- file.path(out_boot, paste0("Resumo_", a, ".csv"))

  if (all(file.exists(c(f1, f2, f3, fR)))) {
    boot_e1_all[[as.character(a)]] <- read_csv(f1, show_col_types = FALSE)
    boot_e2_all[[as.character(a)]] <- read_csv(f2, show_col_types = FALSE)
    boot_e3_all[[as.character(a)]] <- read_csv(f3, show_col_types = FALSE)
    resumo_boot[[as.character(a)]] <- read_csv(fR, show_col_types = FALSE)
    next
  }

  X1 <- as.matrix(dfa[, c("QT_DOC_EXE_trat_log_norm_winsor",
"QT_TEC_TOTAL_trat_log_norm_winsor")])
  Y1 <- as.matrix(dfa[, "QT_MAT_trat_log_norm_winsor", drop = FALSE])
  b1 <- bootstrap_dea_vrs_out(X1, Y1, B = B) %>% bind_cols(dfa %>% select(CO_IES, SG_IES,
Ano))

  X2 <- as.matrix(dfa[, c("QT_MAT_trat_log_norm_winsor",
"QT_ABANDONO_trat_log_norm_winsor")])
  Y2 <- as.matrix(dfa[, "QT_CONC_trat_log_norm_winsor", drop = FALSE])
  b2 <- bootstrap_dea_vrs_out(X2, Y2, B = B) %>% bind_cols(dfa %>% select(CO_IES, SG_IES,
Ano))

  X3 <- as.matrix(dfa[, "QT_CONC_trat_log_norm_winsor", drop = FALSE])
  Y3 <- as.matrix(dfa[, c("SOMATORIO_SERVICOS_trat_log_norm_winsor",
"SOMATORIO_ACERVO_trat_log_norm_winsor")])
  b3 <- bootstrap_dea_vrs_out(X3, Y3, B = B) %>% bind_cols(dfa %>% select(CO_IES, SG_IES,
Ano))

  boot_e1_all[[as.character(a)]] <- b1
  boot_e2_all[[as.character(a)]] <- b2
  boot_e3_all[[as.character(a)]] <- b3

  res_a <- bind_rows(
    b1 %>% summarise(Media_Obs = mean(Eff_Obs), Media_BC = mean(Eff_BC), Dif_Media =
Media_BC - Media_Obs) %>% mutate(Ano = a, Etapa = "Etapa 1"),
    b2 %>% summarise(Media_Obs = mean(Eff_Obs), Media_BC = mean(Eff_BC), Dif_Media =
Media_BC - Media_Obs) %>% mutate(Ano = a, Etapa = "Etapa 2"),
    b3 %>% summarise(Media_Obs = mean(Eff_Obs), Media_BC = mean(Eff_BC), Dif_Media =
Media_BC - Media_Obs) %>% mutate(Ano = a, Etapa = "Etapa 3")
  )

  resumo_boot[[as.character(a)]] <- res_a

  write_csv(b1, f1)

```

```

write_csv(b2, f2)
write_csv(b3, f3)
write_csv(res_a, fR)
}
boot_final_e1 <- bind_rows(boot_e1_all)
boot_final_e2 <- bind_rows(boot_e2_all)
boot_final_e3 <- bind_rows(boot_e3_all)
tabela_resumo <- bind_rows(resumo_boot)
write_xlsx_list(
  list(
    "Boot_Etapa1" = boot_final_e1,
    "Boot_Etapa2" = boot_final_e2,
    "Boot_Etapa3" = boot_final_e3,
    "Resumo" = tabela_resumo
  ),
  file.path(out_boot, "NDEA_BootstrapManual_TodosAnos.xlsx")
)
df_mapa_unique <- df %>%
  select(DMU = CO_IES, SG_IES, TP_REDE) %>%
  distinct(DMU, .keep_all = TRUE) %>%
  mutate(Rede = ifelse(TP_REDE == 1, "Pública", "Privada"))
Data_final_classico <- Data_reload %>%
  filter(RTS == "vrs") %>%
  left_join(df_mapa_unique, by = "DMU") %>%
  select(Ano, DMU, SG_IES, Rede, E1 = etapa1, E2 = etapa2, E3 = etapa3, EG)
out_tabs <- dir_create(out_ndea, "10_Tabelas_Finais")
writexl::write_xlsx(Data_final_classico, file.path(out_tabs, "Tabela_NDEA_Classico.xlsx"))
boot_e1_sel <- boot_final_e1 %>%
  transmute(Ano, DMU = as.character(CO_IES), E1 = Eff_BC)
boot_e2_sel <- boot_final_e2 %>%
  transmute(Ano, DMU = as.character(CO_IES), E2 = Eff_BC)
boot_e3_sel <- boot_final_e3 %>%
  transmute(Ano, DMU = as.character(CO_IES), E3 = Eff_BC)
df_mapa_unique <- df %>%
  transmute(DMU = as.character(CO_IES), SG_IES, TP_REDE) %>%
  distinct(DMU, .keep_all = TRUE) %>%
  mutate(Rede = ifelse(TP_REDE == 1, "Pública", "Privada"))
boot_final <- boot_e1_sel %>%
  inner_join(boot_e2_sel, by = c("Ano", "DMU")) %>%
  inner_join(boot_e3_sel, by = c("Ano", "DMU")) %>%
  mutate(EG = E1 * E2 * E3) %>%
  left_join(df_mapa_unique, by = "DMU") %>%
  select(Ano, DMU, SG_IES, Rede, E1, E2, E3, EG)
writexl::write_xlsx(boot_final, file.path(out_tabs, "Tabela_NDEA_Bootstrap.xlsx"))
# =====
# 2.21. GRÁFICOS DE BOOTSTRAP (ano alvo)
# =====
out_boot_plot <- dir_create(out_boot, "Graficos")
ano_alvo <- max(anos)

b1 <- boot_final_e1 %>% filter(Ano == ano_alvo)
b2 <- boot_final_e2 %>% filter(Ano == ano_alvo)
b3 <- boot_final_e3 %>% filter(Ano == ano_alvo)
png(file.path(out_boot_plot, paste0("Boxplots_", ano_alvo, ".png")), width = 1400, height = 500)
par(mfrow = c(1, 3), mar = c(5, 5, 3, 2))
boxplot(b1$Eff_Obs, b1$Eff_BC, names = c("Determinístico", "Bootstrap"), main = "Etapa 1", ylab =
"Eficiência (VRS, OO)", ylim = c(0, 1))
boxplot(b2$Eff_Obs, b2$Eff_BC, names = c("Determinístico", "Bootstrap"), main = "Etapa 2", ylab =
"Eficiência (VRS, OO)", ylim = c(0, 1))

```

```

boxplot(b3$Eff_Obs, b3$Eff_BC, names = c("Determinístico", "Bootstrap"), main = "Etapa 3", ylab =
"Eficiência (VRS, OO)", ylim = c(0, 1))
dev.off()
plot_disp <- function(df_stage, titulo, arquivo) {
  png(file.path(out_boot_plot, arquivo), width = 1400, height = 700)
  par(mar = c(5, 5, 3, 2))
  n <- nrow(df_stage)
  plot(df_stage$Eff_Obs, ylim = c(0, 1.1), xlab = "DMUs", ylab = "Eficiência (VRS, OO)", main = titulo,
pch = 1)
  points(df_stage$Eff_BC, pch = 5)
  for (i in 1:n) {
    segments(i, df_stage$CI_Lower[i], i, df_stage$CI_Upper[i], col = "gray40")
  }
  legend("bottomright", bty = "n", legend = c("Obs (1/θ)", "BC (1/θ)", "IC 95%"), pch = c(1, 5, NA), lty =
c(NA, NA, 1), col = c("black", "black", "gray40"))
  grid(col = "lightgray", lty = "dotted")
  dev.off()
}
plot_disp(b1, sprintf("Dispersão — Etapa 1 (%s)", ano_alvo), sprintf("Disp_E1_%s.png", ano_alvo))
plot_disp(b2, sprintf("Dispersão — Etapa 2 (%s)", ano_alvo), sprintf("Disp_E2_%s.png", ano_alvo))
plot_disp(b3, sprintf("Dispersão — Etapa 3 (%s)", ano_alvo), sprintf("Disp_E3_%s.png", ano_alvo))
# =====
# 3.2. CARREGAR BASE TRATADA (EG + VARIÁVEIS AMBIENTAIS)
# =====
dados_ml <- readxl::read_xlsx(
  file.path(base_dir, "dados_ml.xlsx")
) %>%
  janitor::clean_names()
dados_ml <- dados_ml %>%
  mutate(
    centroide = as.character(centroide),
    centroide = stringr::str_replace_all(centroide, "POINT\\((\\|\\)", ""),
    centroide = stringr::str_trim(centroide)
  ) %>%
  tidyr::separate(
    col = centroide,
    into = c("lon", "lat"),
    sep = " ",
    remove = FALSE,
    extra = "drop",
    fill = "right"
  ) %>%
  mutate(
    lon = as.numeric(lon),
    lat = as.numeric(lat),
    tp_rede = ifelse(tp_rede == 1, 0, 1)
  )
# =====
# 3.3. RECEITA DE PRÉ-PROCESSAMENTO
# =====
# Objetivo metodológico:
# Esta etapa define a receita de pré-processamento dos dados
# utilizada nos modelos de aprendizado de máquina aplicados no
# segundo estágio da análise empírica.
#
# Procedimento:
# A variável dependente do modelo é a eficiência global (eg), e as
# variáveis explicativas correspondem aos fatores ambientais e
# institucionais considerados na dissertação. A receita específica:
# (i) identificação das variáveis de controle e identificação;

```

```

# (ii) remoção de preditores sem variabilidade;
# (iii) imputação de valores ausentes para variáveis numéricas;
# (iv) centralização das variáveis numéricas; e
# (v) padronização das variáveis numéricas.
#
# Resultado:
# O objeto resultante (rec) reúne o conjunto de transformações
# necessárias para preparar os dados de entrada dos modelos
# preditivos, assegurando consistência no tratamento das variáveis
# e comparabilidade entre os preditores utilizados no segundo
# estágio da dissertação.
#
rec <- recipes::recipe(
  eg ~ dens_mun + pibpc_mun + ano_mec + lat + lon +
  tp_rede + in_capital + dummy_fies + dm_u + ano,
  data = dados_ml
) %>%
  recipes::update_role(any_of(c("dm_u", "ano")), new_role = "id") %>%
  recipes::step_zv(all_predictors()) %>%
  recipes::step_impute_median(all_numeric_predictors()) %>%
  recipes::step_center(all_numeric_predictors()) %>%
  recipes::step_scale(all_numeric_predictors())
# =====
# 3.12. EXPORTAÇÃO FINAL
# =====
# Objetivo metodológico:
# Esta etapa realiza a consolidação e exportação dos resultados
# produzidos pelos modelos de aprendizado de máquina aplicados no
# segundo estágio da análise empírica.
#
# Procedimento:
# São reunidos, em um único arquivo de saída, os principais objetos
# gerados ao longo da etapa de modelagem, incluindo:
# (i) o ranking comparativo de desempenho dos modelos estimados;
# (ii) as predições obtidas a partir dos modelos selecionados; e
# (iii) a importância relativa das variáveis explicativas.
#
# Resultado:
# Os resultados são exportados para um arquivo Excel contendo
# múltiplas abas, permitindo a organização e posterior análise
# comparativa dos modelos utilizados para explicar a eficiência
# global das instituições analisadas.
#
out_dir_ml <- dir_create(base_dir, "12_ML_Segundo_Estagio")
objs <- list()
if (exists("rank_tbl")) objs$ranking_modelos <- rank_tbl
if (exists("preds")) objs$predicoes <- preds
if (exists("varimp_tbl")) objs$importancia_variaveis <- varimp_tbl
writexl::write_xlsx(
  objs,
  path = file.path(out_dir_ml, "ML_Segundo_Estagio_EG.xlsx")
)
cat("\n✓ Resultados finais salvos em:", file.path(out_dir_ml, "ML_Segundo_Estagio_EG.xlsx"), "\n")
# =====
# 3.12.b. BETA REGRESSION
# =====
# Objetivo metodológico:
# Esta etapa estima um modelo de regressão beta com o objetivo de
# investigar a associação entre a eficiência global das instituições
# (eg) e um conjunto de variáveis ambientais e institucionais

```

```

# selecionadas no segundo estágio da análise.
#
# Fundamentação teórica:
# A regressão beta é apropriada para variáveis dependentes contínuas
# limitadas ao intervalo (0,1), como é o caso dos escores de
# eficiência utilizados nesta dissertação. Essa abordagem permite
# modelar adequadamente a média condicional da eficiência em função
# de características observáveis do ambiente institucional e
# territorial, preservando a natureza fracionária da variável
# resposta.
#
# Procedimento:
# Inicialmente, verifica-se a existência da variável dummy_fies na
# base de dados. Em seguida, estima-se o modelo de regressão beta
# utilizando como variável dependente a eficiência global (eg) e,
# como covariáveis, os indicadores ambientais, espaciais e
# institucionais selecionados.
#
# Resultado:
# O modelo estimado permite avaliar a direção e a significância da
# associação entre os fatores ambientais e o nível de eficiência
# institucional, fornecendo evidências complementares à análise DEA/
# NDEA desenvolvida na dissertação.
#
if (!"dummy_fies" %in% names(dados_ml)) {
  dados_ml <- dados_ml %>%
    mutate(dummy_fies = ifelse(qt_mat_fies > 0, 1, 0))
}
fit_beta <- betareg::betareg(
  eg ~ dens_mun + pibpc_mun + ano_mec + lat + lon + tp_rede +
  in_capital + dummy_fies,
  data = dados_ml
)
summary(fit_beta)
# =====
# 3.12.b.1. RESUMO DESCRITIVO POR REDE, FIES E ANO
# =====
resumo_eff <- dados_ml %>%
  group_by(tp_rede, dummy_fies, ano) %>%
  summarise(
    media_eff = mean(eg, na.rm = TRUE),
    mediana_eff = median(eg, na.rm = TRUE),
    desvio_eff = sd(eg, na.rm = TRUE),
    n = n(),
    .groups = "drop"
  ) %>%
  mutate(
    tp_rede = ifelse(tp_rede == 0, "Pública", "Privada"),
    grupo_fies = ifelse(dummy_fies == 1, "Com FIES", "Sem FIES")
  ) %>%
  arrange(tp_rede, grupo_fies, ano)
print(resumo_eff)
# =====
# 4. MALMQUIST (IMQT) por estágio (deaR)
# =====
# Objetivo metodológico:
# Esta etapa estima o Índice de Produtividade de Malmquist por
# estágio da rede produtiva, com o objetivo de avaliar a dinâmica
# intertemporal da produtividade das instituições ao longo do
# período analisado.

```

```

#
# Fundamentação teórica:
# O Índice de Malmquist permite decompor a variação da produtividade
# em dois componentes principais: mudança de eficiência técnica
# (catch-up effect) e mudança tecnológica (frontier-shift effect).
# No contexto desta dissertação, sua aplicação por estágio permite
# identificar em qual parte do processo produtivo educacional ocorreram
# os principais avanços ou perdas de desempenho ao longo do tempo.
#
# Procedimento:
# Inicialmente, é construída uma base balanceada contendo apenas as
# DMUs presentes em todos os anos do recorte analisado, condição
# necessária para a estimação consistente do índice. Em seguida,
# aplica-se a rotina do pacote deaR para cálculo do índice de
# Malmquist em cada um dos três estágios do modelo, utilizando
# orientação a outputs e retornos variáveis de escala.
#
# Resultado:
# O procedimento gera, para cada DMU, ano e estágio:
# (i) o Índice de Malmquist (IMQT),
# (ii) a variação tecnológica (TC),
# (iii) a mudança de eficiência técnica pura (PECH),
# (iv) a mudança de eficiência de escala (SECH), e
# (v) a mudança total de eficiência (EC).
#
# Os resultados são consolidados em uma única base e exportados para
# posterior análise da evolução temporal da produtividade das
# instituições por estágio da rede.
#
out_malm <- dir_create(out_ndea, "11_IMQT_Resultados")
df_malm <- bd_DEA %>%
  filter(Ano %in% anos) %>%
  mutate(DMU = as.character(CO_IES)) %>%
  group_by(DMU) %>%
  filter(n_distinct(Ano) == length(anos)) %>%
  ungroup() %>%
  arrange(Ano, DMU) %>%
  select(
    DMU, Ano, SG_IES, TP_REDE,
    QT_DOC_EXE_trat_log_norm_winsor,
    QT_TEC_TOTAL_trat_log_norm_winsor,
    QT_MAT_trat_log_norm_winsor,
    QT_ABANDONO_trat_log_norm_winsor,
    QT_CONC_trat_log_norm_winsor,
    SOMATORIO_SERVICOS_trat_log_norm_winsor,
    SOMATORIO_ACERVO_trat_log_norm_winsor
  )
# Checagem obrigatória do painel balanceado
contagem_malm <- df_malm %>%
  count(Ano)
print(contagem_malm)
# Garantir que todos os anos têm o mesmo número de DMUs
stopifnot(length(unique(contagem_malm$n)) == 1)
# Garantir unicidade DMU-ano
stopifnot(!any(duplicated(df_malm %>% select(DMU, Ano))))
rodar_malmquist_estagio <- function(df, inputs, outputs, nome_estagio) {
  malm_data <- deaR::make_malmquist(
    datadea = df,
    percol = 2,
    dmus = 1,

```

```

inputs = inputs,
outputs = outputs,
arrangement = "vertical"
)

res <- deaR::malmquist_index(
  malm_data,
  orientation = "oo",
  rts = "vrs",
  type1 = "cont",
  type2 = "fgnz"
)

mlong <- function(mat, var) {
  as.data.frame(mat) %>%
  tibble::rownames_to_column("Ano") %>%
  pivot_longer(-Ano, names_to = "DMU", values_to = var)
}

mi <- mlong(res$mi, "IMQT")
tc <- mlong(res$tc, "TC")
pe <- mlong(res$pech, "PECH")
se <- mlong(res$sech, "SECH")

mi %>%
left_join(tc, by = c("Ano", "DMU")) %>%
left_join(pe, by = c("Ano", "DMU")) %>%
left_join(se, by = c("Ano", "DMU")) %>%
mutate(
  Ano = as.integer(Ano),
  EC = PECH * SECH,
  Estagio = nome_estagio
) %>%
left_join(
  df %>%
  group_by(DMU) %>%
  summarise(
    SG_IES = first(SG_IES),
    TP_REDE = first(TP_REDE),
    .groups = "drop"
  ),
  by = "DMU"
) %>%
mutate(
  TP_REDE = case_when(
    TP_REDE == 1 ~ "Pública",
    TP_REDE == 2 ~ "Privada",
    TRUE ~ as.character(TP_REDE)
  )
) %>%
relocate(SG_IES, TP_REDE, .after = DMU)
}

malm_e1 <- rodar_malmquist_estagio(
  df_malm,
  inputs = c(5, 6),
  outputs = c(7),
  nome_estagio = "Estágio 1"
)

malm_e2 <- rodar_malmquist_estagio(
  df_malm,

```

```
inputs = c(7, 8),
outputs = c(9),
nome_estagio = "Estágio 2"
)
malm_e3 <- rodar_malmquist_estagio(
  df_malm,
  inputs = c(9),
  outputs = c(10, 11),
  nome_estagio = "Estágio 3"
)
malm_todos <- bind_rows(malm_e1, malm_e2, malm_e3) %>%
  arrange(Estagio, DMU, Ano)
writexl::write_xlsx(
  malm_todos,
  file.path(out_malm, "Indice_Malmquist_IMQT_Por_Estagio.xlsx")
)
```

APÊNDICE F – ESTUDOS EMPÍRICOS SOBRE EFICIÊNCIA EM IES

Autor(es)	Contexto / Amostra	Metodologia	Principais Resultados
Internacional			
Charnes, Cooper e Rhodes (1978)	Introdução da DEA	DEA (modelo CCR)	Trabalho seminal que inaugura a análise de eficiência com múltiplos inputs e outputs.
Johnes (2006)	100 IES do Reino Unido	DEA + bootstrap	Alto nível de eficiência; bootstrap permitiu definir intervalos de confiança precisos.
Lee (2011)	37 universidades australianas	DEA-bootstrap + regressão (Simar e Wilson, 2007)	Determinantes positivos: localização e proporção de professores seniores; número de alunos teve efeito negativo.
Parteka e Wolszczak-Derlacz (2013)	266 IES de 7 países europeus (2001–2005)	DEA-bootstrap + Malmquist	Aumento médio anual de 4,5% na produtividade; mudanças de eficiência prevaleceram sobre progresso tecnológico.
Veiderpass e Mckelvey (2016)	944 IES em 17 países europeus	DEA + bootstrap	Eficiência associada ao tamanho da IES e à intensidade de pesquisa; diferenças relevantes entre DEA clássica e bootstrap.
Guccio, Martorana e Monaco (2016)	Universidades italianas (2000–2010)	DEA + bootstrap (Simar e Wilson, 1998)	Reformas de Bolonha aumentaram eficiência; evidências de convergência, mas com lacunas regionais persistentes.
Barra e Zotti (2016)	Universidade de Salerno (Itália)	DEA-VRS + bootstrap + Malmquist	Eficiência dependente da especificação; proxies de qualidade reduzem resultados.
Andersson <i>et al.</i> (2017)	35 IES na Suécia (2005–2008)	DEA + bootstrap	Ineficiência média de 12%; produtividade cresceu 1,7% ao ano.
Navarro-Chavez, Gómez Monge e Torres Hernández (2017)	32 universidades públicas mexicanas (2012)	DEA + bootstrap	Eficiência média de 0,77 (VRS) e 0,71 (CRS); nenhuma universidade atingiu eficiência máxima.
Wolszczak-Derlacz (2017)	348 universidades europeias e 152 americanas	DEA semiparamétrica (2 estágios)	Alta ineficiência técnica; doações e economias de escala aumentam eficiência; variabilidade entre países.
Lehmann <i>et al.</i> (2018)	133 universidades (Alemanha e Itália, 2006–2011)	DEA + regressão (Simar e Wilson, 1998)	Estrutura de governança impacta eficiência; italianas reagiram melhor à crise de 2008.
Brzezicki (2020)	59 universidades públicas e 34 privadas da Polônia	DEA-BCC + regressão truncada	Públicas mais eficientes em formandos; privadas melhores em salários de egressos.
Civera <i>et al.</i> (2020)	72 universidades alemãs e 51 italianas (2004–2013)	DEA-VRS + bootstrap + regressão	Alemanha inferior em eficiência, mas superior em pesquisa; efeitos positivos e negativos da Iniciativa de Excelência.
Martínez-Campillo e Fernandez-Santos (2020)	47 universidades públicas espanholas (2002–2013)	DEA-bootstrap + regressão truncada	Universidades tornaram-se mais eficientes durante a crise de 2008; localização e idade da IES influenciam desempenho.

Autor(es)	Contexto / Amostra	Metodologia	Principais Resultados
Agasisti <i>et al.</i> (2021)	449 universidades russas (2012–2015)	DEA-bootstrap (2 estágios)	Eficiência estatisticamente ligada ao crescimento econômico regional.
Nazarko e Šaparauskas (2014)	Universidades públicas europeias	DEA	Defendem inclusão de variáveis qualitativas na mensuração.
Arbula e Bleich (2024)	30 países europeus (2013–2020)	DEA Window	Identificou variações temporais significativas; destacou relevância de análises dinâmicas.
Guironnet e Peypoch (2018)	Universidades dos EUA Nacional (Brasil)	DEA + análise geográfica	Eficiência associada a fatores regionais e geográficos.
Marinho, Resende e Façanha (1997)	Universidades federais brasileiras	DEA	Primeiras estimativas de eficiência das IFES.
Façanha e Marinho (1999)	IFES	DEA	Relacionam eficiência ao modelo de financiamento, sob perspectiva de contratos e incentivos.
Belloni (2000)	IFES	DEA	Indicadores de eficiência a partir da qualidade da graduação e pós-graduação.
Façanha e Marinho (2001)	IES públicas e privadas	DEA	Estimativas comparativas entre os dois segmentos.
Costa <i>et al.</i> (2012)	49 IFES (2004–2008)	DEA-SBM + Malmquist + jackstrap	Identificação de outliers; eficiência superior em instituições com forte atuação em pós-graduação e pesquisa.
Siqueira (2015)	Universidades federais (2007–2012)	DEA + Malmquist	Ganhos modestos de eficiência técnica, atribuídos à escala (Reuni).
Santos, Silva Júnior e Nunes (2022)	56 universidades federais (2012–2018)	DEA-BCC (output-oriented)	Eficiência média de 94,5%; crescimento de produtividade de 2,2% (catch-up).
Silva e Rosa (2022)	37 universidades federais (2014–2016)	DEA	Mais da metade eficiente; necessidade de ampliar critérios de avaliação.
Almeida <i>et al.</i> (2018)	Departamentos de Economia (2009–2012)	DEA-S + bootstrap	Queda de 20% na eficiência técnica; condicionada a fatores socioeconômicos dos alunos.
Souza (2021)	FIES (2000–2020)	Estudo qualitativo	Expansão do acesso ao ensino privado, mas com custos fiscais e desafios de sustentabilidade.
Pontuschka (2016)	FIES (2010–2015)	Avaliação de impacto	Mais de 1,1 milhão de contratos; cerca de R\$ 14 bilhões mobilizados.
Becker e Mendonça (2021)	FIES	Estudo empírico	Redução da evasão, mas aumento do tempo de conclusão e inadimplência.

Fonte: Elaboração própria com base na base de dados Web of Science.